Ed Tech Rapid Cycle Evaluation Coach

HANDBOOK

Table of Contents

Ed Tech Rapid Cycle Evaluation Coach

Rapid Cycle Evaluation Overview

INTRODUCTION

In an ideal world, as you decide whether to continue using an educational technology or to purchase a new license, you would base your decision on how well the product meets your needs and improves outcomes. Mathematica Policy Research, the U.S. Department of Education, and SRI International are supporting quick-turnaround evaluations to help schools test and evaluate these products. We call this process a Rapid Cycle Evaluation (RCE). The findings generated through this process can help you make decisions by providing you with evidence regarding how well an educational technology works.

We have built a free, web-based toolkit—the Ed Tech RCE Coach—that will guide you through the steps of conducting an evaluation. As you move through the Ed Tech RCE Coach, the tools will help you to design and conduct your own product evaluations and pilot tests.

The RCE Coach can help you answer several important questions:

- Does a specific educational technology program lead to the student outcomes you want to see?
- Should you keep paying for a software tool you're already using?
- Does a software tool support more effective teacher professional development?
- How should I set up an effective pilot?

EXAMPLES

Several districts have already begun conducting RCEs:

- **A school district in Mississippi** is conducting two cycles of evaluations of a personalized English language arts technology used with struggling readers. For both cycles, the district is conducting **randomized controlled trials**[1] to evaluate the effectiveness of the technology. First, the district investigated the effect of using the technology in a summer school program. It compared the standard curriculum without the technology to a curriculum with the technology during a four-week summer reading program. The results of the summer evaluation will inform the district's use of the technology during the full school year, with the

[1] A **randomized controlled trial**, or RCT, is a type of study in which a lottery or other random process determines the groups that will be compared. With a well-implemented RCT, you can be confident that the educational technology you're studying caused any differences in outcomes between the two groups.

aim of informing further expansion of the program. This is a perfect example of the cycle part of Rapid Cycle Evaluations.

- **A charter school system in Texas** was interested in how a reading technology used by individual students as part of a supplemental reading program affected student reading achievement. A limited number of schools had already implemented the technology, so the school system used a **matched comparison design**[2] to evaluate the effects. The charter school system investigated the effects by grade, and the results informed its decision about whether to implement the technology in other schools throughout the district.
- **A district in Illinois** is interested in how several technologies that have been purchased (covering both math and English language arts) contribute to student achievement. Currently, the district plans to use a matched comparison design to evaluate the effectiveness of the various technologies.
- **A district in Colorado** wishes to test how a professional development support product can aid implementation of a new literacy curriculum. Teachers will be invited to participate in the training program, but only a subgroup will use the software. The district will analyze outcomes using teacher surveys to determine whether use of the product is associated with better results.

[2] When it is not possible to create groups based on chance (as an RCT does), a **matched comparison design** attempts to create groups that are as similar as possible. It does this by using data on characteristics of students, teachers, or schools to create groups that are similar on key factors that could be related to outcomes.

Guide to Using the Coach

What You'll Need to Get Started

This document outlines some basic requirements for running your own rapid cycle evaluation and the steps you will take through the Coach. Along the way you might have to bring in other members of your team (such as a data analyst) to help.

Before you can begin crafting an evaluation within the Coach you will need the following:

- An educational technology to test (can already be in use, or not)
  - If you don't have a technology, our Choose a Technology guide can help you select one
- A population using or willing to use the technology
  - Within the population of potential users you will have to be able to create two similar groups, one using the technology and one not using the technology
- A program for collecting and organizing data, such as Microsoft Excel

REQUIRED INFORMATION

Exhibit 1 summarizes the information you will need to complete various steps within the evaluation. The earlier you begin collecting and thinking about this information the easier it will be to complete your evaluation.

Exhibit 1. Information needed for the evaluation

Step	Information needed
The basics	• Who the technology users are (or will be) • What outcomes you're interested in
Determine your approach	• Whether the technology has been implemented • How you will sort potential users into similar groups
Craft your research question	• What outcome you are targeting • Who you are trying to affect
Think about how you will use the results	• Cost of the technology • Idea of what success looks like
Summarize context	• Basic information about the technology and its implementation • Details about the educational environment

REQUIRED DATA

The earlier you begin collecting and cleaning your data the better. Reviewing our data guide, Prepare Your Data for Analysis, will help acquaint you with the required format for the data you will upload to the Coach and how to achieve that format. Depending on the type of evaluation you are conducting, you will have to use data at different points. It is likely that you won't be able to gather all of your information at once, but the earlier you can get to a completed data set, the easier your analysis will be. Note that it is important that all of your data are recorded using individual identifiers. These are unique codes for each participant. The identifiers—student, teacher, or school identifiers—are what will enable you to combine (merge) data sets without using any personally identifiable information (PII). The RCE Coach does not require personally identifiable student information to use the tool and does not create evaluation reports or briefs that contain PII. When you follow the instructions provided by the tool carefully, your use of the tool will comply with requirements of the Federal Educational Records Privacy Act. A list of the data you will need follows (for more details, refer to the data guide linked above).

- Outcome data (required)
- Outcome or outcome-related data from before use of the technology (required for certain designs)
- List of who is and is not using the technology (required)
- Background characteristics (important for establishing a comparison group)
- Technology usage data (required for certain research questions)

If you are not managing the data for the evaluation, the planning stage is an important time to communicate with the person responsible for data management. You should discuss the data required for your specific evaluation, and determine how and when those data will be collected.

REQUIRED KNOWLEDGE AND SKILLS

Throughout the course of your evaluation, you and/or someone on your team will have to possess the following skills and competencies:

- Ability to answer questions about the technology being tested and the desired outcomes
- Understanding of how to collect or extract data from different sources
- Proficiency with a data management system

Ed Tech Rapid Cycle Evaluation Coach

ICON DIRECTORY

Guide: These documents are meant to assist you in completing your evaluation by providing you with educational material about the evaluation process.

Interactive tool: These tools require you to input information relevant to your evaluation.

Statistical tools: These tools analyze the data you input to determine who is in your treatment and comparison groups and whether your technology is moving the needle.

Guide: Choose a Technology

Educational technology can help to achieve the goals you've set for your school or district. But selecting the technology that best fits your needs can be challenging. The RCE Coach is targeted to those who have already selected or implemented a technology. If you are looking for a technology to implement, this guide will help you with that process and you can return to the Coach when you are ready to plan a pilot of that technology and start your evaluation.

FOUR STEPS FOR SELECTING THE RIGHT TECHNOLOGY

1. Conduct a needs assessment.
2. Discover what technology might fit your needs.
3. Vet potential technologies.
4. Begin initial conversations with technology providers.

1. CONDUCT A NEEDS ASSESSMENT.

A needs assessment will help you identify gaps (the needs), set priorities, and determine important criteria for solutions—the educational technology you choose. This assessment can take many forms. It can be informal conversations or interviews with various teachers, students, or administrators, or it could be a more formal survey, focus group, or working group. You will have to decide what kind of needs assessment you would like to engage in based on the time and resources you have available. If you feel as though you already know the needs of your school or district and have a sense for how educational technology can help you meet those needs, you might not have to conduct a needs assessment. But if you do choose to conduct a needs assessment, we've included some helpful guidelines:

A. **Determine your knowledge objective(s).** It's important to clearly outline what you want to learn from the assessment you are conducting. Write out some questions you want to answer at the end of the assessment based on the information you collect. This will help you determine who to talk to and what questions to ask them. For example, if you want to know what features might be most effective in a new technology, you might want to talk to students about the types of virtual tools and games they already use. If you want to know if you should target specific subgroups of students with the technology, you might want to talk to teachers about which students need the most help. Make sure to write out each of these knowledge objectives and refer back to them throughout the process of designing and conducting your assessment.

B. **Choose who you want to participate in your assessment.** Before you begin conducting your assessment, you should determine who you want to hear from. Perhaps your objectives require

input from a wide variety of sources, or maybe you want to focus only on teachers, administrators, or students. Determine who you'll need to speak to and invite them to participate in your assessment.

C. **Develop your questions.** When you know who will participate, develop the questions that you will ask them to reach your knowledge objectives.

D. **Conduct your assessment.** Now you can begin your assessment. You'll want to record the responses in some way. For an informal assessment this could just mean taking notes during or after your conversations. For a more formal survey, this might mean recording responses in a systematic way.

E. **Analyze your data and draw conclusions.** After you've spoken to everyone you need to, look back at your data (notes or survey responses) for trends and themes. Have multiple participants made similar comments? Are there needs that people already agree on? These are the things you'll want to consider as you start looking for the right educational technology.

2. DISCOVER WHAT TECHNOLOGY MIGHT FIT YOUR NEEDS.

Here are some questions to keep in mind while you look at the various educational technologies that are available:

- What problem are you trying to solve or what opportunity do you wish to pursue?
- Who should be involved in the selection process?
- What are the technical requirements?
- How much training is required and how will it be delivered?
- What's the intended implementation approach?
- What kind of data or reporting tools do you need?

A number of websites collect information like this, and more, to help you select the right technology. Another good starting point is the What Works Clearinghouse, which can help you find technologies that have been effective in other settings.

3. VET POTENTIAL TECHNOLOGIES.

In addition to existing requirements and discoveries from the needs assessment, here is a checklist of questions to consider when vetting technologies:

☐ Does the implementation model you have in mind match what the tool is designed to do?
- Small groups versus classrooms
- Remediation versus general practice versus enrichment
- Self-directed versus system-directed
- In-class versus independent use
- Adaptable versus assignable

- ☐ Product fit
  - What problem do you wish to solve, and how does the technology address this problem?
  - Does the technology address the desired outcomes for learning, productivity, and so on?
  - Does the theory of change or learning underlying the technology match your approach?
  - Are student-focused technologies aligned with the standards you care about?
- ☐ Product design
  - Do intended users find the technology engaging?
  - Is it accessible for all target users?
  - How easy is the technology to use?
- ☐ Implementation issues
  - How easy is the technology to use?
  - Do teachers, students, or other intended users have to undertake some preparatory training? Does the developer provide such training?
  - Is the technology compatible with your school system's existing technology?
  - For technologies meant to be used out of school, do intended users have access to the necessary devices and connectivity?
- ☐ Cost and time needed for implementation
  - What are the license costs?
  - Will the developer provide a free demonstration period? (This is strongly encouraged!)
  - What costs are associated with implementing the technology? Are any additional equipment purchases involved?
  - What costs are associated with administrative and teacher time for training, reviewing the technology, and incorporating it into the curriculum?
  - Evidence of effectiveness
  - Is there strong evidence of the effectiveness of the technology?
  - Is the evidence of effectiveness specifically for students and settings like yours?
- ☐ System data availability
  - What learning and usage indicators does the technology track?
  - Do the indicators tracked by the technology provide the information you need to determine whether it is effective in achieving your goals?

4. BEGIN INITIAL CONVERSATIONS WITH TECHNOLOGY PROVIDERS.

When you've narrowed down the number of potential educational technologies, you should begin talking to providers and developers about your needs and specific requirements. Our guide to conversations with product developers can help you start those conversations.

Ed Tech Rapid Cycle Evaluation Coach

Guide: The Basics

The Basics enables you to enter simple information about your educational technology. This is the information that the Coach will use throughout your evaluation to personalize and refine your experience. Here is a guide to the questions included in The Basics; you can use this document for additional guidance on how to answer these foundational questions.

Question	Explanation	Example
Have you selected an educational technology for your evaluation?	The Coach is designed to help you evaluate an educational technology. To begin an evaluation and answer all subsequent questions in the Coach, you will have to know what you are evaluating. The Coach is not designed to help you select or compare potential technologies to evaluate, but you can refer to our brief guide—Choose a technology—if you have not yet selected a technology to evaluate.	You may select only yes or no for this question.
What is the name of the educational technology?	This refers just to the name of your educational technology. By answering this question, the Coach will be able to ask about your specific technology as you move through your evaluation. This will also be used in your findings brief.	Example: eZumi Learning
What is the outcome you hope to change by using the technology?	This is a broader question about your goals for the technology. This will help the Coach modify future questions to reflect the outcome you are interested in. In a later step you will provide more detail on how you will measure this outcome and the specific changes you want to see. You can find examples for how to categorize your goals after this table.	Example: student academic achievement
Who are or will be your technology users?	This question refers to the people directly engaging with the technology. In some cases, both teachers and students will engage with the technology. You should select the users for whom you are interested in measuring outcomes or, if you think it is important to identify all users, you can select other and type in "students and teachers."	Example: students

HOW TO SELECT AN OUTCOME

Outcomes are knowledge, skills, attitudes, or other desired benefits attained as a result of an activity (for example, engaging with the educational technology). We encourage you to meet with your team and determine the outcome you are most interested in measuring during your evaluation. When you have selected an outcome, you will have to categorize it in one of four categories. Some examples follow to help you choose which category to select. If none of these categories cover your outcome, you can select other and enter a description.

Student academic achievement	Choose this option if you are using a technology to improve student performance on an academic measure. This mostly refers to outcomes you can evaluate using a subject test of some sort. Select this if your technology targets: • Math performance • Reading or literacy skills • Science understanding
Student non-academic achievement	Choose this option if you are using a technology to improve student outcomes on a non-academic measure. This mostly refers to outcomes you can evaluate by observing behaviors. Select this if your technology targets: • Chronic absenteeism • Graduation rates • College application or matriculation rates
Teacher performance	Choose this option if you are using a technology to improve teacher outcomes. This mostly refers to technologies aimed at improving outcomes for teacher performance evaluations. Select this if your technology targets: • Teacher professional development • Lesson planning
Parental engagement	Choose this option if you are using a technology to change the way parents engage with their children's education. Select this if your technology targets: • Attendance at parent–teacher conferences • Reading at home

Guide: Determine Your Approach

There are multiple ways to conduct an RCE in your school or district. Before you can begin working through your evaluation, we'll help you find the approach that works best for you. The decision tree provided here will help you understand the pathway that suits your unique context. Please note that by answering the questions in this step of the RCE Coach, the Coach can determine this for you.

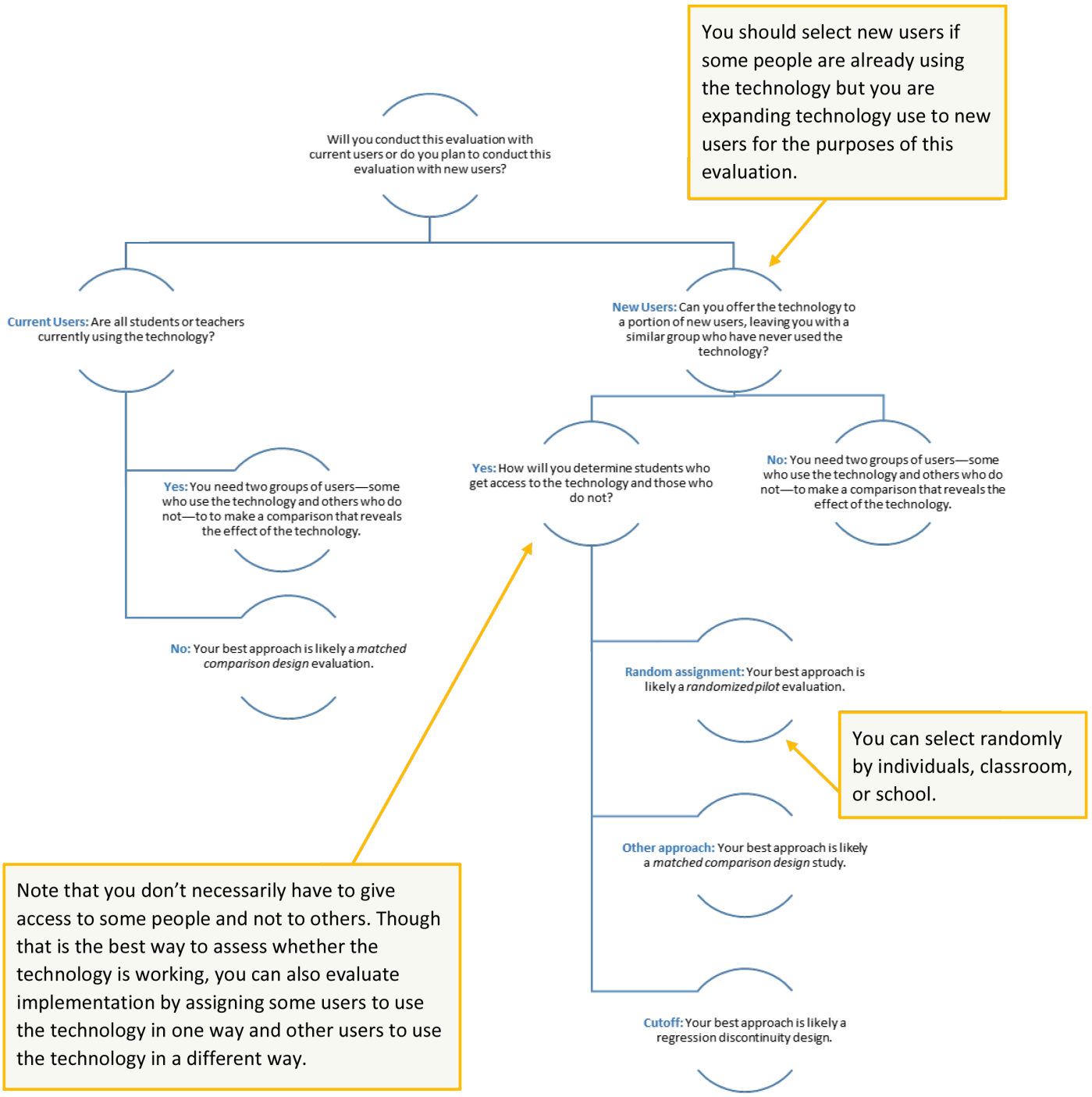
The RCE Coach is a powerful tool that helps you move quickly to make decisions based on rigorous evidence—but you'll have to gather some data to get there. It is never too early to start thinking about how you will do that, how the approach you choose will affect those efforts, and whether you'll have to bring in any additional team members to help with that process.

Note that there are four potential approaches (two of which are currently supported):

1. Randomized pilot
2. Matched comparison design
3. Regression discontinuity design
4. None of the above

Please follow the decision tree in Exhibit 1 to determine your recommended approach. The decision tree will also help you to understand what changes you can make to ensure you can continue with either a randomized pilot or matched comparison design.

Exhibit 1. Decision tree



CHOOSING AN ASSIGNMENT METHOD

The method you choose will affect how confident you can be that your technology causes the outcomes you observe and how your findings will apply to a broader population of students.

Random assignment—recommended

Random assignment is similar to putting the names of your students in a hat and randomly picking some of the names. If you put in 10 names and select 5, those you pick would get to pilot (or try) the technology, and those not selected would continue with your existing program or curriculum. In this scenario, each individual would have a 50 percent chance of being chosen. The Coach can conduct this process for you using a digital hat and show you the similarities between the groups. This process can also work with teachers, classes, or even schools.

Advantages

- The resulting treatment and comparison groups will, on average, be similar in both observed (for example, gender) and unobserved (for example, motivation) characteristics, giving you the most confidence that any differences in outcomes between the groups are due to using the technology.
- Random assignment is a fair way to decide who gets to use a technology first, because everyone has the same chance of being chosen.
- Findings are relevant for the overall population with which you are working (that is, everyone in the hat).

Disadvantage

- Some students get to use the technology earlier than others, which might be a problem for some stakeholders.

Cutoff

Using a cutoff means selecting who uses the technology based on where students rank in relation to an existing threshold. For example, students just below a low-score threshold are assigned to use the technology. We assume that students close to a cutoff are very similar; therefore, you can compare those just below the cutoff with those just above it.

Advantage

- A cutoff enables you to compare similar users without using random assignment.

Disadvantages

- Findings apply only to students near the cutoff, not to the whole population.
- This procedure might be less fair than random assignment. With this method, scoring 1 point below the cutoff gives you a 100 percent chance of using the technology, whereas scoring 1

point above the cutoff means you have a 0 percent chance of using the technology, even though both groups of students might be equally in need of it.
- This method is not currently supported by the RCE Coach. Contact us if you'd like to discuss testing this method.

Other approach

Assignment to the treatment and comparison groups takes place in a nonrandom manner based on specific criteria. For example, the classrooms assigned to the pilot or treatment group are those with access to computers or the schools that will pilot the technology are chosen based on a competition or application process.

Advantages

- This process might align better with how you've assigned technology use in the past.
- It might be logistically easier, particularly if you're selecting based on who has better access to computers or the Internet.

Disadvantage

- Stakeholders might not consider this process to be fair. The group that meets the criteria might already have advantages over other students.
- It might not align with need. Those who need the technology more or would gain more from using it might be less likely to receive access than if you used random assignment.
- It's not always possible to create a good comparison group. If the Coach fails to create similar groups of users and nonusers, you will not be able to determine whether the educational technology is moving the needle. It's possible that groups appear similar, but unobserved differences, not included in your data, are driving outcomes.

NOTE: Across all of these methods, you do not have to create or compare groups at the individual level. If there are logistical barriers to working with individuals, you can create groups at higher levels, such as classrooms or schools. However, at higher levels you will need more study participants to be confident in your results.

Guide: Who is Using Your Technology—and How?

INTRODUCTION

Usage data (also sometimes called system or log data) can be a good way to find out **who** is using education technology and **how**. This information can shed light on **why** a product is or isn't having the desired effect. Pages 16 through 22 of this guide will walk you through the steps of getting and using usage data in your evaluation. Appendix A, pages 75-87 provide you with fictional case studies, providing additional examples on how to use usage data as well as a glossary of terms, and data templates.

There is no one format for usage data, as the information captured by a system varies depending on the technology's intended purpose. Here is one example of usage data:

Exhibit 1. Usage data

Schools ID, class IDs, or grades are useful for analyzing groups of students

Student ID	School ID	Grade	Total # of sessions	Total # of minutes	Average minutes per session	Percentage of quizzes attempted	Average quiz score
12345	123	2	10	125	12.5	66	80
12346	123	2	10	130	13.0	71	78
12347	123	2	14	245	17.5	50	64
12348	123	3	1	10	10.0	90	86
12349	123	3	2	17	8.5	84	95
12350	123	3	0	0	NA	75	80
12351	123	3	1	7	7.0	69	55

NA = not applicable.

Each row represents an individual student

These columns provide insight into how much a student engaged with the system

These columns shed some light on the student's progress and performance

This example provides a useful summary of technology use and can address some basic questions about which students used the technology and how often they used it.

NOTE: You will need more detailed data to explore questions about:

- Usage over time, such as whether students started out using the technology a lot but their use then dropped off after a few months
- Students' performance on specific topics, such as a unit on geometry or reading comprehension

STEP 1. DETERMINE THE INFORMATION YOU NEED

To get helpful usage data from developers, you will have to communicate exactly what information—that is, fields in the data set—you want extracted from the technology. Figuring out what data you want starts with determining what questions you wish to answer. For example:

- **Is the educational technology being used as intended?** To answer this, you might want to know which students (or other users) are logging into the system, how frequently they log in, and how long they typically spend logged in.
- **Are students showing signs of learning in the system?** To answer this, you might want to determine whether students attempt embedded assessments and how they score. Consider, too, how well the assessments in the system align with the learning objectives.
- **Are students 'gaming' the system?** This might be the case if you see students making lots of attempts at assessments, or if they spend very little time using instructional materials.
- **Are teachers using the technology? I**nformation about when and how often teachers log in to the system could tell you whether they are using the technology to monitor students' progress or for other purposes.

NOTE: By combining usage data with administrative data on students' background characteristics or outcomes, you can explore important questions that can't be addressed with usage data alone. For example, does usage behavior vary for different types of students? Is more time spent in the educational technology associated with better outcomes?

STEP 2: GET USAGE DATA FROM THE EDUCATIONAL TECHNOLOGY PROVIDER

If the educational technology has a feature to download reports, start by seeing what information is readily available. Do these reports provide the data you need, or will you have to contact the educational technology provider for a custom data file?

Ed Tech Rapid Cycle Evaluation Coach

When requesting a custom data file, consider the following:

- **Do you want data for individual students?** Student-level data is necessary for analyzing questions about different types of students or correlations between technology use and external outcomes.
- **Do you want data to be grouped by teacher, class, grade, or school?** Aggregating data by groups can make data sets more manageable. In general, averages are more useful than raw totals. For example, it is more helpful to know the average amount of time spent per student for each class (basically the total time divided by the number of students) than the total amount of time spent.

Here are examples of questions you can address with different levels of data:

If you have …	You can address these types of questions …
Student-level data	• What is the association between use of the technology and students' outcomes? • Do students with different background characteristics (for example, prior knowledge or socioeconomic status) have different outcomes associated with the technology? • Does usage behavior for students vary by their characteristics?
Classroom-level data	• Which classes actively used the technology? • How often and for how long was the technology used across classes? • Did class growth on particular topics correlate with use of corresponding modules in the technology?
School- or district-level data	• How many teachers used the system? • Over what time period was the system used? • What was the average session duration across the school or district? • What features or types of content were used most frequently?

TIP: Getting student-level data with class and school IDs provides you the flexibility to aggregate at different levels to address different questions. If you get data that are already aggregated by class or school, you will have less flexibility to analyze the data in different ways.

ADDITIONAL CONSIDERATIONS

- **How will you match usage data from the technology with district data on students' outcomes and backgrounds?** Try to obtain data files with student identifiers that match with other data sets, such as district IDs.
- **What type of file do you want?** To analyze usage data, you will need reports in formats that allow you to manipulate the data, such as Microsoft Excel or CSV files.
- **Do you want long- or wide-format data?** Long-format data have multiple rows per student, whereas wide-format data have one row per student. Having multiple rows per student can be more manageable if, for example, you want data on students' progress and performance for lots of time periods or topic areas.

Exhibit 2. Example of long-format data

Note that some student IDs appear on multiple rows

Student ID	Class ID	Goal	Average score	% of learning path completed
12345	123	Geometry	90	100
12345	123	Measurement and data	60	30
12345	123	Number and operations	67	50
12346	124	Geometry	85	60
12346	124	Measurement and data	77	55
12347	125	Geometry	89	100
12347	125	Measurement and data	68	40
12347	125	Number and operations	76	50
12347	125	Operations and algebraic thinking	56	70
12348	450	Geometry	92	40

Exhibit 3. Example of wide-format data

Student ID	Class ID	Geometry: average score	Geometry: % of learning path completed	Measurement and data: average score	Measurement and data: % of learning path completed	Number and operations: average score	Number and operations: % learning path completed
12345	123	90	100	60	30		
12346	124	67	50	85	60		
12347	125	89	100	68	40	76	50
12348	450	92	40				

STEP 3: GENERATE DESIRED TABLES

The data files you receive from technology providers might not present information in exactly the format you want. For example, if you receive multiple lines for each student showing performance on different modules, you might wish to aggregate these to generate one row for each student (as in Exhibit 3). Or, you might wish to add additional columns to calculate averages and ranges to address the specific questions you seek to answer.

Exhibit 4. Worked example 1: Is the product being used? How often?

Student ID	Class ID	Total # of sessions	Total # of minutes	Average minutes per session
12345	123	10	125	12.5
12346	123	10	130	13.0
12347	123	14	245	17.5
12348	450	1	10	10.0
12349	450	2	17	8.5
12350	450	0	0	NA
12351	450	1	7	7.0

NA = not available.

With some manipulation, you can produce the following table that addresses the research questions more directly:

Exhibit 5. Average usage by classroom

Classroom	% using at all	% using recommended amount*	Average minutes / session (excludes non-users)	Range of minutes / session (excludes non-users)
All	85%	50%	14.1	7.0-17.5
123	100%	100%	14.3	12.5-17.5
450	75%	25%	8.5	7.0-10.0

Exhibit 6. Worked example 2: How are students performing on assessments within the system?

Student ID	Class ID	Goal	Average score	% of learning path completed
12345	123	Geometry	90	100
12345	123	Measurement and data	60	30
12345	123	Number and operations	67	50
12346	124	Geometry	85	60
12346	124	Measurement and data	77	55
12347	125	Geometry	89	100
12347	125	Measurement and data	68	40
12347	125	Number and operations	76	50
12347	125	Operations and algebraic thinking	56	70
12348	450	Geometry	92	40
12349	450	Geometry	88	32
12350	450	Geometry	71	50

In Class 450, students perform relatively well on tests despite being less likely to complete the learning path. It is possible that the teacher is using the product as an optional supplemental resource.

With some additional manipulation, the following table can be produced that describes performance at the student level:

Exhibit 7. Student-level performance

Students	Average score	% learning path completed
All	76.6	62
12345	72.3	60
12346	81.0	57
12347	72.3	65

The student with the highest average score has completed a relatively low percentage of learning paths. Again, this suggests that the product is being used as a supplemental resource

Alternatively, you can create a table showing class performance on various goals in the system:

Exhibit 8. Class-level performance

Class ID	Goal	Average score	% learning path completed
All		76.6	62
123	Geometry	72.3	60
123	Measurement and data	81.0	57
123	Number and operations	72.3	65

STEP 4: MATCH WITH ADMINISTRATIVE RECORDS

Many evaluations involve analyzing associations between students' use of technology and their performance on external assessments, such as standardized tests. Most study designs also involve comparing students with similar background characteristics, meaning that you should also factor in background data such as age, gender, race and ethnicity, and prior academic achievement.

In these cases, after you have cleaned the usage data and organized it the way you want, you will have to merge the usage data file with administrative records. This is where having common, unique student IDs is important.

Guide: Craft Your Research Question

A carefully crafted research question grounds the development of an RCE in a clear statement of the evaluation's goals. This guide highlights the key components of a well-defined research question and provides examples for each of the four types of research questions that you can answer using the RCE Coach.

INGREDIENTS OF AN EFFECTIVE RESEARCH QUESTION

Exhibit 1 illustrates the key ingredients of a focused research question, or the "**A, B, C**, and **D**s". The blueprint for an effective research question is structured as follows:

Does **A** do **B** among **C** compared with **D**?

Thinking through the details of the intervention; its intended goal; and the target population of students, teachers, or schools at this stage will make the rest of the process much easier. However, you might not know every aspect detailed in this table at this point. If that's the case, the table can help identify critical aspects of the evaluation that you will want to discuss with key stakeholders as soon as possible.

EXPECTED ANSWERS TO A WELL-FORMED RESEARCH QUESTION

A rapid cycle evaluation will provide one of three answers:

1. Yes, **A** is likely to do **B** among **C**, compared with **D**.
2. No, it is not likely that **A** does **B** among **C**, compared with **D**.
3. More data are needed to reach a strong conclusion.

The answer you get could drive important decisions within your school or district. The Coach's guide for Planning Your Next Steps will help you think through how you will use the evidence you uncover through this process. Because a well-designed evaluation aims to provide a clear answer to the precise question it was designed to address, it is important to confirm with evaluation stakeholders that the question addresses a learning objective that can help inform decisions and actions.

Exhibit 1. Components of a well-crafted research question

The blueprint: Does A do B among C compared with D?	
"**A**" is the name of the educational technology you are testing. Consider including how often and for how long the users interact with or are exposed to the program.	**Examples:** • "eZumi Learning" • OR • "12 hours of professional development on the use of TrueStar" • "Six weeks of using a modified version of eZumi three times per week"
"**B**" is the intended effect of using the technology. To identify "B" you will have to determine your outcome of interest (what you hope to change by using the technology), the direction of the intended change (increase or decrease), and the measure you will use to determine the effect (if applicable). An outcome is knowledge, skills, attitudes, or other desired benefits that are attained as a result of an activity (for example, engaging with the educational technology). We encourage you to meet with your team and determine the outcome you are most interested in measuring during your evaluation.	**Examples:** • "Increase achievement on an interim test of math proficiency" • "Increase the number of videos posted in a four-week period" • "Increase minutes of technology use" • "Decrease absences"
"**C**" is the group of people for whom you want to see results. What group of people are you trying to affect? For example, students might be using the technology to increase their own achievement or teachers could be using the technology to increase their students' achievement. In both cases, the answer to Part C would be students.	**Examples:** • "6th-grade students" • "3rd- through 5th-grade teachers" • "Students who were not proficient on last spring's state math test" • "Algebra teachers who have previously used eZumi in their classroom" • "5th-grade English learners"
"**D**" is the group of people you want to compare the group defined in "C" against. This group should be as similar to those identified in C as possible. In an ideal world, the only difference between the two groups would be the intervention. When thinking about this group, consider where the technology is being used. If the technology is being used in only three of five classes, then the other two classes might be the appropriate comparison group. Please note that it could be necessary to look at school- or district-wide data to obtain data for an appropriate comparison group.	**Examples:** • "Similar 5th-grade English learners with no access to eZumi Learning" • "Similar 5th-grade English learners who use Wizlet Math instead of eZumi Learning" • "Similar algebra teachers who received only six hours of professional development on the use of TrueStar"

RESEARCH QUESTION EXAMPLES

The RCE Coach is designed to help answer four key types of research questions. Exhibit 2 lists examples of each type.

Exhibit 2. Examples of research questions, by type

Category of research question	Examples of clear, narrow research questions
Does this technology achieve its intended outcomes?	• Does a semester of daily use of Everest Learning Lab by teachers increase math test scores among middle school students compared with middle school students whose teachers do not use Everest Learning Lab? • Does a semester of daily use of Everest Learning Lab by teachers increase math test scores among students who are below grade level in math achievement compared with similar students whose teachers do not use Everest Learning Lab? • Does TrueStar increase match achievement among 2nd graders compared with similar 2nd graders who are using eZumi?
Does this training help users engage with this technology more and/or better?	• Does providing teachers with two days of training on how to implement a flipped classroom using Everest Learning Lab increase the number of videos posted in a four-week period, compared with teachers who did not receive the training?
Does providing information to [parents, teachers, or students] change behavior?	• Does sending automatically generated, weekly text reminders to parents increase time spent reading in eZumi over the course of a month among elementary school students, compared with similar students whose parents did not receive reminders? • Does an automatically generated text message to parents when a student is absent in homeroom increase students' attendance over the course of a semester, compared with similar students whose parents did not received text messages?
Does this modification to the technology (or how it is implemented) make it work better?	• Does changing the order of the content in eZumi for two months increase performance on benchmark assessments among middle school students, compared with similar students who see the content in the default (or different) order? • Does altering the interface to be more visual and less verbal increase lesson completion rates over the course of three weeks among students, compared with students who see the default interface? • Does using TrueStar in an after-school, peer-guided setting increase students' test scores on a benchmark assessments among English learner students compared to English learner students who use TrueStar in an independent lab-based setting?

Guide: Think About How You Will Use the Results

We'll use this tool to think about how the answer to your research question will inform the decisions you make regarding your technology. The following questions will help you prepare you to think about the relationship between your results and the technology costs and stakes associated with your decision. To complete this tool you will have to consider the following:

What would success look like?

How much risk are you willing to take when making a decision or a recommendation?

CONSIDER: When making decisions, it is important to think about the costs of the technology you're considering and the risks associated with your decision.
For example, you might be trying to decide:

- Whether to keep paying for a software tool you're already using
- Whether to buy licenses for a tool you're piloting on a trial basis
- Whether to change how teachers and/or students use the tool, based on a new approach tested in your evaluation

A. WHAT IS THE COST OF THE EDUCATIONAL TECHNOLOGY?

The cost of a technology will probably be a factor in your decision about how big its effect has to be in order to be a good value. Cost might also influence how certain you want to be that the technology has the desired effect.

The cost of a technology could simply be the cost per user. It could also be the savings realized by replacing one technology with another less expensive technology. Cost can also be measured in something other than dollars, such as teacher preparation time saved.

How much do you pay (or save) per student, teacher, or school to use the educational technology?

Examples

- It costs about $20 per student to use the educational technology.
- It saves about $1,000 per classroom to use the educational technology.
- We're not sure. We believe using this tool will save teachers 20 minutes per day on preparation time.

B. WHAT WOULD SUCCESS LOOK LIKE?

The Coach will calculate the probability that the educational technology has an effect at least as large as the number you choose here. The Coach uses the units of whatever you are using to measure outcomes. For example, if you are looking at student performance on a test, the units would be points.

CONSIDER: This number could be 0 or even negative. A 0 means that you would consider any increase in the outcome a success, as long as it is positive. If the technology saves you a lot of money, you might be willing to consider any change that is greater than a small negative number a success.

The RCE Coach will tell you "There is an X percent probability that the intervention increases the outcome by Y units or more." In this step you select the value for Y.

What is the direction of intended change? (Do you hope to increase or decrease your outcome of interest?)

What measure will you use to test whether there was an effect?

Examples

- Midyear math assessment
- End-of-year state reading assessment

What is the unit used to measure your outcome of interest?

Examples

- Test score points
- Percentage points

By how many units would your outcome have to increase or decrease for you to consider the technology a success?

To answer this question, think about what impact size would be meaningful in your context. If you hope to see an increase in a reading test score, would a 1 point increase, on average, be a meaningful increase? What about 5 points? When deciding what a meaningful change would be, you should consider factors such as the maximum possible points and current performance. A 10-point increase on a test with a maximum possible score of 100 is very different from a 10-point increase on a test with a maximum score of 500. Similarly, if students score 80 out of 100 on average, there is less room for improvement than if they score 50 out of 100 on average, and that might affect the size of the improvement that you would find meaningful.

C. HOW CONFIDENT DO YOU WANT TO BE?

Rarely do we have enough evidence to be nearly certain (close to 100 percent certainty) that an educational technology product produces the desired results. Most of the time, we have to make choices with incomplete evidence. Thinking forward to your results, what probability would you need to be comfortable concluding that the educational technology is moving the needle?

As you think about what the probability threshold would be for you to conclude that the technology had the intended effect, consider the stakes involved. You might want a higher level of certainty if your results will influence a high-stakes decision than if your results will be used for a lower-stakes decision. For example:

- High stakes—if you are deciding whether to implement a district-wide curriculum
- Low stakes—if you are deciding how to use a small amount of class time

Though it might be tempting to set a high probability threshold for any kind of evaluation, it is important to note that the higher the threshold the harder it will be to meet. This might not necessarily be because the technology isn't working, but could be driven by the number of users you are testing the technology with or your unit of assignment. With a very small number of users, you will be unlikely to achieve high confidence even if the product is very effective.

Example

The RCE Coach will tell you "There is an X percent probability that the intervention increases the outcome by Y units or more."

That is, there is a 91.1 percent probability that the intervention increases the outcome by 5 scale score points or more.

In this step, you are selecting the value to which you will compare X. If X is less than your chosen value, you would not be confident that the technology is moving the needle. If X is greater than your chosen value you would be confident that the technology is moving the needle.

That is, if you selected 95 percent in this step, you would conclude that the technology is not moving the needle because 91.1 percent is less than 95 percent.

D. WHAT WILL YOU DO IF …

This RCE will provide you with one of the following three answers to your research question. It is important that you think about what you will do under each scenario.

1. … it is likely that the educational technology has the intended effect?
2. … it is not likely that the educational has the intended effect?
3. … the results are inconclusive?

Examples

If it is likely that the educational technology has the intended effect …

- *We will roll out eZumi to all students in grades 3 through 5.*

If it is not likely that the educational technology has the intended effect …

- *We will not renew our eZumi license for the next school year.*

If the results are inconclusive …

- *We will collect more data to get a higher level of certainty about our results.*

Guide: Work with your Ed Tech Provider

Working with your educational technology provider can be crucial for getting the information you need for a successful evaluation. Providers can be useful at multiple stages throughout the process, from helping you understand best practices for implementation to getting you the usage data you need to understand who is using your technology, and how. This tool will help you start a productive conversation with your provider.

WHY WORK WITH PROVIDERS

Setting up an initial call with the education technology provider is a good way to accomplish the following:

- Learn about the educational technology product
- Learn how the school or district is using the product
- Learn about embedded assessments within the educational technology product
- Determine what data reports are available
- Identify appropriate external outcomes to measure for your evaluation

Consider inviting the following staff members to participate in this conversation:

- Curriculum coordinator
- Data analyst
- Procurement staff member
- Information technology staff member
- Instructional staff member

HOW TO GET THE MOST OUT OF YOUR CONVERSATION

The following steps are a good starting point for planning and beginning your conversation with your provider.

1. **Review the Guide to Usage Data.** Reviewing this guide will help you know what data you need to get from the provider.
2. **Use our sample email message.** You can use this as a template for contacting providers via email to set up a meeting. Change the appropriate fields before sending. You might also want to add any specific or unique questions you have for your provider.

Dear [Ed tech provider contact],

[Name of school/district] plans to conduct an evaluation of [name of product's] use in our schools. As part of the planning for this evaluation, we would like to set up an introductory phone call between your company and district staff who are participating in the evaluation. The purpose of this call is to [discuss how we can obtain system data needed for the evaluation]. It would be ideal to include people at [name of company] who are familiar with the implementation in our district and with the product's data reporting capabilities.

If you are not the right person at [name of company], please forward this message to whomever you think would be best placed to speak with us about these topics. Also, kindly suggest some times that would work on your end for a phone call. We expect this will take about 60 minutes.

Thank you in advance for your cooperation!

Regards,

[name, title and contact info]

3. **Use our sample discussion guide.** This guide can help to organize your initial conversation with developers. Edit it to fit your needs, or tick off the topics you want to be sure to cover.

WHAT DATA TO REQUEST

Our usage data guide (Who used your technology—and how) provides a more comprehensive overview of the data you might want to discuss with your provider. However, a brief summary of this information follows.

- When requesting a system data file, consider the following:
  - Do you want data for individual students?
  - Do you want to group data by teacher, class, grade, or school?

Here are examples of questions you can address with different levels of data:

If you have …	You can address these types of questions …
Student-level data	• What is the association between use of the technology and student outcomes? • Do students with different background characteristics (for example, prior knowledge or socioeconomic status) have different outcomes associated with the technology? • Do usage behaviors vary among students with different characteristics?
Classroom-level data	• Which classes actively used the technology? • How often and for how long was the technology used across classes? • Was class growth on particular topics associated with use of corresponding modules in the technology?
School- or district-level data	• How many teachers used the technology? • Over what time period was the technology used? • What was the average session duration across the school or district? • What features or types of content were used most frequently?

Tip: Getting student-level data with class and school IDs provides you the flexibility to aggregate at different levels to address different questions. If you get data that are already aggregated by class or school, you will have less flexibility to analyze it in different ways.

Additional considerations

- **How will you match system data from the technology with district data on student outcomes and backgrounds?** Try to obtain data files with student identifiers that match with other data sets, such as district IDs.
- **What type of file do you want?** To analyze usage data, you will need reports in formats that allow data to be manipulated, such as Microsoft Excel or CSV files.
- **Do you want long- or wide-format data?** Long-format data have multiple rows per student, whereas wide-format data have one row per student. Having multiple rows per student can be more manageable if, for example, you want data on students' progress and performance for lots of time periods or topic areas.

Ed Tech Rapid Cycle Evaluation Coach

DISCUSSION GUIDE

Use this as a template for structuring your initial conversations with educational technology providers.

☐ Introduce the RCE project: Consider whether you should:

- Explain the purpose of the evaluation. Is your goal to:
  - Pilot a new educational technology product?
  - Evaluate the impact of an existing product?
  - Decide whether to renew a license or expand use of the product?
  - Improve implementation?
- Share initial ideas about the type of evaluation: Will it be backward-looking (that is, examining outcomes for those already using the technology in a certain way) or forward-looking (that is, examining outcomes for new users or those being told to use the technology in a new or specific way)? For forward-looking evaluations, how do you plan to assign access to the technology?
- Describe what you need from the educational technology provider to move forward with the evaluation

☐ [If not already known] Ask the educational technology provider to describe the product and intended benefits.

- Target users: For whom is this technology designed? What is its intended purpose or benefit?
- Underlying learning science and active ingredients: What makes this product unique? What is the theory of change?
- Is the product aligned with your state or district standards?
- What can the developer tell you about optimal usage? For example, how often should the technology be used, for how long, and over what time frame?

☐ [For educational technologies already in use] Ask for a description of how and how much the educational technology product is being used.

- Which schools and grades use the product?
- How many students and teachers use the product?
- Does the developer have any insights about implementation strengths and areas for improvement at your district or school?
- What professional development or training has the developer provided? What does the developer recommend?
- Should any other important implementation characteristics be taken into account?

Ed Tech Rapid Cycle Evaluation Coach

- ☐ Ask the educational technology provider to describe any embedded assessments within the technology.
  - What are they based on? Are they demonstrated to be valid and reliable?
  - Do they correlate with any external measures, such as standardized tests?
  - What are the most important indicators of progress within the technology?
- ☐ Discuss external outcomes to measure and sources of data.
  - What kinds of outcomes is the educational technology intended to affect?
  - How could these outcomes be measured? What data sources could be used to measure their impact? (For example, would standardized test scores be a good way to measure impact?)
  - Over what time frame would you expect to see changes in performance?
- ☐ Determine what data reports are available from the educational technology product.
  - Can any reports be generated automatically?
  - Is it possible to obtain customized data reports?
  - Is a data dictionary (a document that explains the meaning of each field in the usage data) available?
- ☐ Agree on next steps. Consider whether you should:
  - Schedule a product demo?
  - Obtain sample data?
  - Obtain a data dictionary?
  - Schedule a follow-up call?

SOME COMMON DATA TERMS

This short glossary is designed to clarify the various types of data terminologies and associated definitions.

Attempts A record of how many times a student tried to complete a particular assessment or assignment

Clickstream data A record of the sequential order of users' events and mouse clicks in a system, often used to analyze usability of a technology interface

Duration The amount of time elapsed during a session; cumulative duration for a term or academic year can be a useful variable for understanding the overall level of exposure with a technology

Event Any discrete action a user takes within a system, such as submitting responses to an assessment or viewing an instructional video

Learning data Information related to users' interactions with the technology indicating student achievement on formative and summative assessments and/or capturing student behaviors related to learning

Metadata Information related to the content or topic addressed by a particular unit within a program; for example, metadata could be a goal such as "adding fractions"

Performance metrics Information related to users' achievement or progress during system use; for example, a metric could be how many assignments or tasks a student has completed

Reporting function Many systems provide reports summarizing usage, progress, and performance data for a particular group of students (such as a class or school); these reports involve processing raw system log data and are typically downloadable as Microsoft Excel, CSV, or portable document format (PDF) files

Session The user's interactions with a technology during one sitting, often between logging in and logging out

System log data Information related to users' interactions with the technology recorded over time

Time stamps Information about when an event took place (for example, time and date for completing an action)

Guide: Summarize Context

When you share the results from your study, the information you provide through the Coach will help your colleagues make sense of the educational technology implementation and the results. Study context is also important for understanding how the evaluation applies to other environments.

By describing the context using the questions outlined below, your report will include important contextual features about the specific technology used by your study. This will also make it easier to search for other reports on similar technologies or similar settings.

WHAT IS THE EDUCATIONAL TECHNOLOGY?

These questions are meant to provide additional information about the basic elements of the technology you are using.

Exhibit 1. Characteristics of your technology

Questions	Explanation	Example
What is the evaluation period?	This is the date that users start engaging with the technology, and the date on which you'll be measuring the outcome(s) of interest.	Students begin using eZumi on September 10, 2017, and outcomes are measured on December 1, 2017.
What type of program is the educational technology?	This is the area that the program is meant to affect or be used in.	eZumi is a curriculum.
What is the purpose of the educational technology?	This is a general description of the technology you are using and its intended effect.	The technology is a software program designed to improve reading achievement among under-performing students.
What are the key components of the educational technology?	These are the different features of the technology you use, such as how it presents information to students or tracks progress.	The technology features a screening test and adaptive diagnostic testing, as well as use of video content to engage students in skill-building tasks and a data interface that enables teachers to track students' progress.
What guidelines, if any, does the developer provide regarding how the educational technology should be used or implemented?	This refers to any directions you receive from the provider regarding how often the technology should be used, when it should be used, who should use it, or other details meant to assist you in using the technology.	The developer recommends students spend 90 minutes per week using eZumi Learning.

HOW IS THE EDUCATIONAL TECHNOLOGY BEING USED?

These questions provide additional information about how the technology is being used and what it is being used for. Your answers here should reflect how you plan to implement the technology. You will have an opportunity to document and explain any implementation issues or deviations from this plan later in your evaluation.

Exhibit 2. Characteristics of technology implementation

Questions	Explanation	Example
How often and for how long are users supposed to use the educational technology?	This refers to how the users are meant to engage with the technology, in terms of frequency and duration of use, throughout the study.	Twice per week for 40 minutes per session
How was the educational technology delivered for your evaluation?	This refers to whether you the technology was used independently by individuals, used in small groups, delivered to a whole class, or provided to a whole school.	Individually
What grade level used the educational technology for your evaluation?	This refers to the grade level of the students or teachers using the technology. Select all that apply.	K, 1, and 2
For your evaluation, in what type of classroom was the educational technology being used?	This refers to either a general or inclusive classroom setting.	General
What outcome do you hope to change by using the educational technology?	This question is generally answered in "The Basics." It refers to the main category your outcome of interest falls into, such as student academic achievement (test scores), student nonacademic achievement (attendance), teacher performance, and parent engagement.	Student academic achievement
In which of the following outcome areas will you look for differences between users and nonusers of the educational technology?	These categories are meant to provide a narrower category for the outcome of interest.	Literacy

WHAT IS THE EDUCATIONAL SETTING?

Different technologies work well in some settings and not so well in others. The Coach provides some guiding questions to help you provide relevant information, and space to provide any other contextual information you believe is relevant. You might not have answers to all of these questions. Answer them as best you can with the information you have. This information will help you think about how your findings apply to a wider population within your school or district.

Exhibit 3. Characteristics of the educational setting

Questions	Explanation	Example
Tell us about your school district	Here you should select the type of school the technology is being implemented in, input the number of students in the district overall, choose the appropriate geographic setting, and select the state where the school district is located.	Public/3,000/urban/Massachusetts
Tell us about the students in your district	Here you should input demographic information about the students in your school district overall.	51% female, 25% English learners
Tell us about any other contextual considerations	Include features of the setting that you think might influence how well you would expect the technology to work. These might have to do with the available technology resources (such as computer labs, laptops for students, and so on); student population; or presence of other technology initiatives at schools.	Students had difficulty logging into their accounts for the first two weeks of class. About half of the teachers got frustrated and used other products for the rest of the term.

Guide to Using Randomized Pilot Approach

Overview: Creating Groups with Random Assignment

If you select pilot participants randomly (by chance), you will create two groups that are similar on observed and unobserved characteristics. When random assignment is well implemented, you can be confident that any differences in outcomes are due to the educational technology you are testing. Because of this, random assignment is considered the best choice for evaluations of effectiveness and should be used whenever possible.

THE EVALUATION CHALLENGE

You want to test whether a technology is effective, but it is impossible to simultaneously observe what happens when an individual uses a technology and doesn't use that same technology. If you introduce a technology and watch what happens, you might notice improvements, for example, in student test scores. However, you cannot assume that the technology caused improved student outcomes. Many other factors (including regular classroom teaching, other programs, student maturation, and so on) could have contributed to the increases.

To overcome these challenges, it is important to compare a group of technology users to a group of nonusers, on the assumption that the only real difference between them is whether they are using the technology. However, comparing technology users and nonusers brings an additional set of challenges. When we make comparisons without trying to ensure similarities between groups, it is possible that those who use the technology differ in any number of ways from those who do not use the technology. For example, hard workers might be more likely to try a new technology, but they also perform better on tests. This might cause you to confuse the effect of the technology with the effect of working hard (because either could cause the technology users to outperform nonusers). Those differences can make an ineffective technology look effective, or vice versa.

RANDOMIZED PILOT

Solution

The best test of whether your technology works is to randomly select some of the potential users to pilot the technology and others to continue classroom practice as usual. When the groups are assigned by chance, you can be confident that you are comparing apples to apples—that the two groups are the same in every way except the technology use. Then, if you see differences in outcomes (such as student achievement scores) you can be convinced that the new technology is moving the needle.

Ed Tech Rapid Cycle Evaluation Coach

How it works

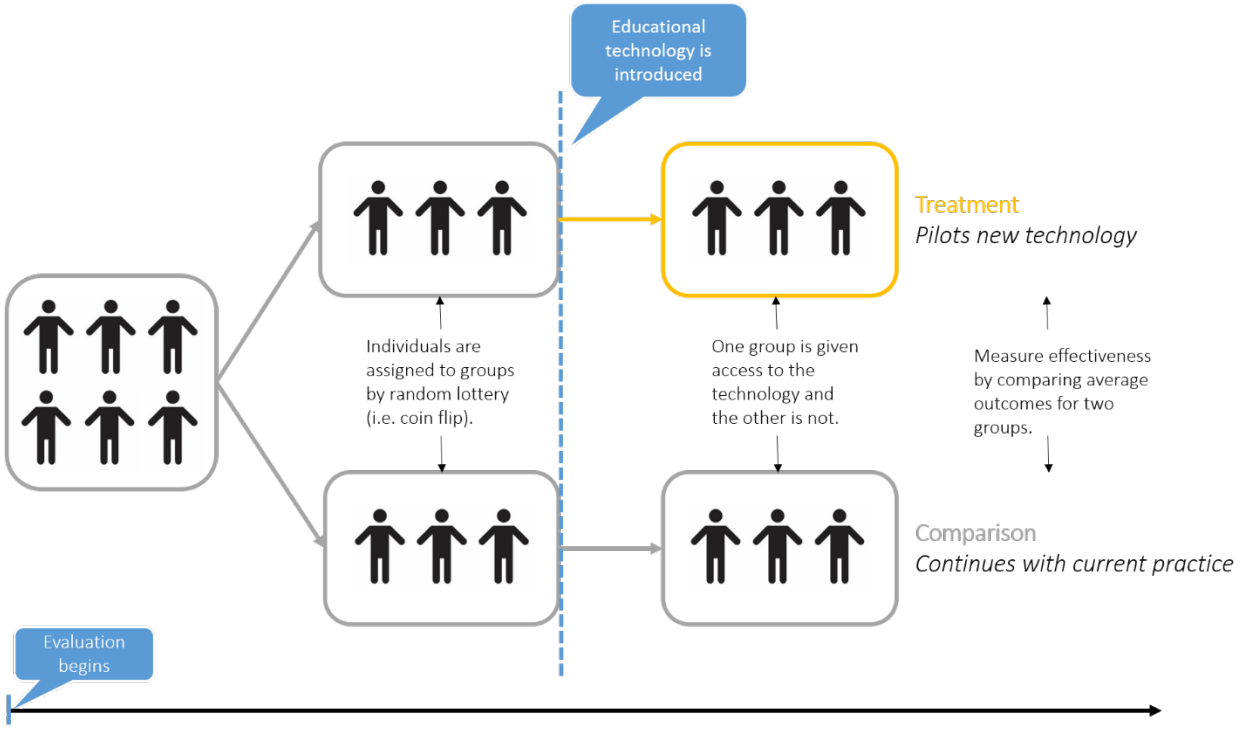
Suppose you want to determine the impact of the reading technology U-Read on 5th-grade reading test scores in your school. To know if U-Read is having an effect on end-of-year reading scores, you would like to be able to observe your users in a parallel universe. In this universe the parallel nonusers are exactly the same as the users, but have no access to U-Read. If this parallel group scored lower on the end-of-year reading test, you could conclude that U-Read moved the needle for users.

With random assignment

You can randomly assign students at the beginning of the year, using a method equivalent to a coin flip, to one of two groups: users or nonusers. You have two groups that are the same on observed and unobserved characteristics on average. Although they aren't really parallel users, the only difference between the groups in the aggregate is whether they use the technology.

Therefore, you can conclude that any differences in achievement are due to the technology, not other factors.

Exhibit 1. Random assignment



BASELINE CHARACTERISTICS

An important question to ask is "Was the random assignment successful?" If random assignment was successful, the two groups will appear similar on measured background characteristics, such as test score results from previous years or demographic characteristics. A common way to assess this is to compare the average values of the groups' background characteristics. We can quantify the difference between the two groups (users and nonusers) using a measure called an **effect size**[1]. The RCE Coach's random assignment dashboard automatically calculates the effect size for any variables you specify. The Coach's random assignment dashboard will run and rerun its randomization until you have baseline equivalence on the characteristics you selected.

The final analysis also uses background characteristics, after you implement your technology and collect outcome data. It is necessary to include background characteristics in your analysis of the results if a lot of participants drop out of the study. The dropout rate in an evaluation is called attrition. For instance, if all students remain in the technology user group but 25 percent drop out of the comparison group, the two groups might no longer be similar. This attrition can occur for many reasons; for example, students might be absent on the day the test is administered or move to a different district or state, or teachers could decide not to participate in the study after random assignment. Accounting for background characteristics can rebalance the two groups, if the initial differences were not too large. However, it is a good idea to include some background characteristics in your analysis even if your evaluation went as planned. (To learn more about attrition, see the U.S. Department of Education's What Works Clearinghouse overview of attrition standards.)

[1] An **effect size** measures different characteristics using the same yardstick. It is calculated by dividing the difference in means between the two groups by the standard deviation of the entire sample.

Guide: Random Assignment

The best test of whether your technology works is to randomly select some of your potential users to pilot the technology (treatment group) while others continue their typical classroom practice (comparison group) and compare outcomes between the two groups. Using chance or random assignment to create the groups increases the likelihood that the new technology causes the differences in outcomes.

HOW ARE YOU ROLLING OUT THE TECHNOLOGY?

Before you can randomly assign users and nonusers, the Coach will need some basic information about how you are implementing the technology.

1. **Who is using the technology?** If you didn't answer this question in The Basics, you will have to provide an answer here. This answer should reflect the type of user who is directly engaging with the technology.
2. **Do you want to randomly assign individuals or groups of individuals (such as classes or schools)?** You do not have to randomly select students at the individual level. It might be more feasible to select by classroom or by school. However, your ability to learn whether the technology is effective will be stronger if you can choose at the individual level. Randomly assigning individuals reduces the likelihood that factors tied to a specific teacher or school cause the outcomes you observe, which might be the case if you assign users at a higher level.

   If you assign by groups, how will you group individuals? Determine whether you are assigning based on class, school, or another factor.
3. **Are there limits to the number of individuals who can access the technology?** Do you have any resource constraints (such as computer or software availability, limited licenses, or time constraints) that would affect the size of your treatment group? If you can offer the technology to only a few individuals, your evaluation will be stronger if your treatment and comparison group sizes are similar (though they need not be equal).

Note: If there are no limits to the number of users you can include, you should try to include as many as possible. The more users you can compare with nonusers, the more confident you can be that any differences in outcomes can be attributed to the technology and not some other factor unique to one group. A difference observed across 100 individuals is much more compelling than a difference between two individuals.

UPLOADING YOUR DATA

Before you upload your data, refer to our guide, Prepare Your Data for Random Assignment, to ensure that the file you upload meets the necessary requirements for the Coach to randomize your participants. Your data file should have one row for each one of your participants (or classes or schools if you're assigning by groups) and one column for each data field (for example, test scores and background characteristics).

Identifying variables

After you upload your data, you will have to tell the Coach which variable names correspond with key pieces of information.

- **Select the variable that contains the anonymous ID that uniquely identifies each individual, class, or school**. Be sure to create a key or a file that links the original names or IDs to the new anonymous IDs. You will need this key to know who was assigned to each group and to add outcome data at the end of the pilot period.
- **Select your pre-test measure variable.** If available, the Coach will use the pre-test measure and background characteristics to check the equivalence or balance of the treatment and comparison groups. If you have a small sample, the Coach might have to randomize multiple times to create balanced groups.
- **Select the variables for the characteristics that should be balanced between the treatment and control groups.** These are characteristics that would make you skeptical of your results if they were found to be unbalanced between groups.

Note: You do not necessarily have to include every available characteristic. You should include only characteristics that you suspect influence the outcome in which you are interested. For example, if your outcome of interest is students' attendance, you might not have to include information about the students' teacher, such as years of experience, as that information is unlikely to affect students' attendance.

Prepare Your Data for Randomization

REQUIRED

At minimum, you must include a list of users with an individual identifier (ID) for each individual, class, or school that will be randomly assigned.

Option 1: If you are assigning individuals, your data set should include an ID for each student or teacher.

AnonStudentID
159508
694677
807588
482489
555123
124226
232721
834305
490514
573401
275321
288475

Option 2a: If you are assigning groups (classes or schools), your data set should include student and class or school IDs.

AnonStudentID	SchoolID
159508	100
694677	100
807588	100
482489	100
555123	100
124226	200
232721	200
834305	200
490514	300
573401	300
275321	300
288475	300

Option 2b: If you are assigning groups, but you do not yet have individual information or won't need it for your final analysis, your data set should include only class or school IDs.

SchoolID
100
101
102
103
104
201
202
203
301
302
303
304

Each row represents a single observation (student, teacher, school, and so on). Each observation must have an individual identifier. This is a unique code for each participant that will be assigned to the technology user group or the comparison group. The identifiers will enable you to determine who will use the technology, so participants can be notified. They will also enable you to combine (merge) data sets.

RECOMMENDED

The Coach recommends that you include in your data set pre-test data and background characteristics. You will use this information to make sure that the randomly assigned treatment and comparison groups are similar before you introduce the technology. If you include these additional variables, your data set will look like this:

Missing data is replaced with NA so that the Coach doesn't include a value for that observation.

Categorical variables must be converted to binary variables with two possible values: 1 instead of yes and 0 instead of no.

AnonStudentID	SchoolID	Grade	Fall_score	Female	EL	Low_SES	Black	White	Asian	Other
159508	100	3	320	1	1	1	1	0	0	0
694677	100	3	450	0	0	0	0	1	0	0
807588	100	4	NA	1	0	0	1	0	0	0
482489	100	4	410	0	1	0	0	0	1	0
555123	100	4	534	1	1	1	0	0	1	0
124226	200	3	604	0	0	0	1	0	0	0
232721	200	3	378	1	0	1	0	0	0	1
834305	200	4	NA	1	1	1	0	0	1	0
490514	300	3	380	0	1	0	1	0	0	0
573401	300	3	468	0	0	1	0	0	0	1
275321	300	4	523	1	0	0	0	0	1	0
288475	300	3	375	0	0	0	0	1	0	0

SES = socioeconomic status.

Each row represents a single observation (student, teacher, school, and so on).

Each variable has its own column

NOTE: You can work with data in a number of different programs. If you do not have access to statistical software, you can use Microsoft Excel to prepare your data. In the rest of this guide, we have included some tips that will help you manage your data in Excel.

STEP 1. IDENTIFYING DATA SOURCES

You should use several types of data to create this initial data set. A checklist and description of the data follows. Examples at the end of this guide show what each data set should look like.

- **List of users who will be randomly assigned (required).** Compile a list of all potential technology users. For example, if you plan to randomly select students to use the technology, you will need an identifier for every student who might be assigned to either the technology user group or the comparison group. If the students are in different schools or different classrooms, you should also include an identifier for the classroom or school. If you plan to randomly select teachers or schools, you will need an identifier for each teacher or each school. (Remember, larger sample sizes are better.)
- **Pre-test data (recommended).** If available, you should include an outcome measure from before the intervention (such as an assessment from the beginning of the school year). The Coach will double-check that the two groups are equivalent before giving you the final list of assignments. If the groups are not balanced, the Coach will rerandomize the list of users until balance is achieved.
- **Background characteristics (recommended).** Background characteristics provide data on observable traits for each participant. These could include gender, ethnicity, individualized education program status, English Learner (EL) status, socioeconomic status (SES), and more. If you include background traits in your data set, the Coach can make sure your intervention and comparison groups are well balanced before introducing the technology.

CAUTION: Some of these characteristics, such as EL status, can change over time. It is preferable to measure and record all background characteristics before introducing the educational technology.

STEP 2. PROCESSING YOUR DATA

When you have identified the data elements and data sources, the second step is to combine all data elements into one **tidy data**[1] set and prepare the variables that will be used for analysis. (We explain how to do this below.) We recommend generating tidy data sets not only because doing so is a requirement to use the RCE Coach and most statistical software packages for analyses, but because a tidy data set is easy to manipulate, model, and visualize. The data set at the beginning of this guide is an example of a tidy data set. This section will take you through a series of questions that will help you create your own.

A. Is each observation a row and each variable a column?

[1] Having **tidy data** means that you've used a standardized way to structure your data set. Specifically: Each variable forms a column; each observation forms a row. If you want to learn more about tidy data you can refer to Wickham, H. "Tidy Data." Journal of Statistical Software, vol. 59. No. 10, 2014, pp. 1–23. doi:http://dx.doi.org/10.18637/jss.v059.i10

Exhibit 1. Example observation

AnonStudentID	Treatment	Fall_score	Gender
159508	1	320	F
694677	0	450	M
807588	0	999	F
482489	1	410	M

NO: Reorganize your data set so that each row represents one observation. Each variable you are interested in should be its own column.

YES: Continue on to B.

B. Do you have one data set that contains all of the variables you will need?

NO: You will have to merge your existing data sets into one complete data set. This will be easy to do using the unique identifiers. If you are using Excel to manage your data, you can do this using a VLOOKUP function.

CAUTION: Some observations could be present in some data sets but not in others. Therefore, when merging these data sets you might introduce some missing data. For example, a student in a data set consisting of test scores might not exist in another data set and therefore could have missing data for other variables (such as background characteristics) after combining the data sets.

YES: Continue on to C.

C. Are all of the categorical variables that will be used for your analysis numeric?

NOTE: If you are using a statistical software package, you will want to make sure that all of these variables are recognized as numeric values and not string or character values.

NO: Convert all of your categorical variables, or variables that include names or labels, into numbers. This might mean you have to change a variable into a binary, or dummy, variable. A dummy variable uses 1 to indicate yes or that a condition was met and 0 to indicate no or that a condition was not met.

For example, if your background characteristics include gender as a variable, you might have male or female, or M or F, in each cell of that column. Instead, you should change the variable from Gender to Female, and change each cell that indicates the participant is a female to 1, and each cell that indicates the participant is male to 0. You can do this for every variable that is non-numeric.

Exhibit 2. Example categorical variables

Student ID	Gender
159508	F
694677	M
807588	F

Student ID	Female
159508	1
694677	0
807588	1

NOTE: If your categorical variable contains more than two options, such as Race—where the options are (1) Asian, (2) Black, (3) White, and (4) Other—you will have to create a binary or dummy variable for each option. For example, Asian would be one column (with 0 representing non-Asians and 1 representing Asians), Black would be a second column, White would be a third column, and Other would be a fourth column (Exhibit 6).

YES: Continue on to D.

D. Are all missing data coded consistently in your data set?

NO: If you have merged data sets and data are missing, make sure you are consistently coding that as NA (not available) to ensure that the Coach can analyze your data. You want to be extra careful to make sure missing data have not been given a numerical designation, such as 0 or 999. These values will get incorporated into the analysis. If you are using your own statistical software, such as SAS or Stata, it will be helpful to code the missing data as a period (.)

YES: Congratulations! You have a tidy data set!

Step 3. Checking the quality of your data

After constructing your data file and converting your variables, the final step is to check the quality of your data. You can run the following checks to identify potential data issues that warrant additional investigation:

Check the minimum and maximum values of variables. This check can help to identify extremely low or high values that are outliers in your distribution or that signal a special missing code that must be converted to a missing value. You might want to check with someone who is familiar with the data to confirm the value range makes sense.

NOTE: If you are working in Excel, you can use MIN and MAX functions to easily find these values.

Consider the impact of missing data. The Coach, and some statistical software packages, will automatically drop observations that contain missing data. You should try to understand why data are missing and how excluding students with incomplete data will affect your results.

NOTE: If you are working in Excel, you can sort and filter your data to view missing values. To determine exactly how many values are missing for a single variable you can use the COUNTIF function; to determine how many observations have at least one missing value, you can use a nested COUNTIF with OR function.

Example data sets

Exhibit 3. Data set 1a: List of participants

Student_name	AnonStudentID	School	SchoolID
Guillermo Gonzalez	159508	Alan Elementary	100
Robert Rice	694677	Alan Elementary	100
Sophia Smith	807588	Alan Elementary	100
Patricia Pacheco	482489	Alan Elementary	100

You will need identifiers for each participant. If you are randomizing groups of participants (classes or schools), you must also include a group ID.

Exhibit 4. Data set 1b: List of participants without personally identifiable information

Eliminate personally identifiable information, such as student and school names, from the data that you upload to the Coach.

AnonStudentID	SchoolID
159508	100
694677	100
807588	100
482489	100

Exhibit 5. Data set 2: Test scores

AnonStudentID	Fall_score
159508	320
694677	450
807588	NA
482489	NA

These are missing values.

Exhibit 6. Background characteristics (with non-numeric and numeric categorical variables)

AnonStudentID	SchoolID	Gender	EL status	SES	Race
159508	100	F	EL	Low	Black
694677	100	M	Not EL	Medium	White
807588	100	F	Not EL	High	Black
482489	100	M	EL	High	Asian
555123	100	F	EL	Low	Asian
124226	200	M	Not EL	Medium	Black
232721	200	F	Not EL	Low	Other
834305	200	F	EL	Missing	Missing

EL = English learner; SES = socioeconomic status.

Data set 3a: Background characteristics (with non-numeric categorical variables)

Data set 3b: Background characteristics (with numeric categorical variables)

AnonStudentID	SchoolID	Female	EL	Low_SES	Black	White	Asian	Other
159508	100	1	1	1	1	0	0	0
694677	100	0	0	0	0	1	0	0
807588	100	1	0	0	1	0	0	0
482489	100	0	1	0	0	0	1	0
555123	100	1	1	1	0	0	1	0
124226	200	0	0	0	1	0	0	0
232721	200	1	0	1	0	0	0	1
834305	200	1	1	NA	NA	NA	NA	NA

EL = English learner; SES = socioeconomic status.

The Coach requires that all categorical variables are converted to binary or dummy variables with values of 0 and 1. In the case of SES, you can choose to group medium and high SES together (1 = low and 0 = medium or high) if you are only interested in the effect of low SES on your outcome of interest.

Note: Research shows that students of higher SES often have an academic advantage over students of lower SES because of differences in early education access, home enrichment, levels of stress, food access, and many other factors. Therefore, if possible, it is good to check that the two randomly assigned groups are balanced on SES. However, some schools or districts might not have access to SES measures or might not be able to use them in this evaluation because of privacy concerns.

Prepare Your Data for Analysis: Randomized Pilot

INTRODUCTION

To answer your research question you will need data from multiple sources. If you need guidance on processing your data, this document will take you through best practices. At the end you will have a data file that will look like this and be ready to upload to the Coach:

Exhibit 1. Summary table

Missing data are replaced with NA so that the Coach doesn't include a value for that observation

Categorical variables must be converted to binary variables with two possible values: 1 instead of yes and 0 instead of no.

AnonStudentID	SchoolID	Grade	Treatment	Fall_ Score	Spring_ Score	Female	EL	Low_SES	Black	White	Asian	Other
159508	100	3	1	320	35	1	1	1	1	0	0	0
694677	100	3	0	450	52	0	0	0	0	1	0	0
807588	100	4	0	NA	37	1	0	0	1	0	0	0
482489	100	4	1	410	89	0	1	0	0	0	1	0
834305	200	4	1	NA	58	1	1	NA	NA	NA	NA	NA
490514	300	3	0	380	NA	NA	NA	NA	NA	NA	NA	NA
573401	300	3	0	468	45	NA	NA	NA	NA	NA	NA	NA
275321	300	4	1	523	78	NA	NA	NA	NA	NA	NA	NA
288475	300	3	0	375	37	NA	NA	NA	NA	NA	NA	NA

EL = English learner; SES = socioeconomic status.

Each row represents a single observation (student, teacher, school, and so on).

Each variable has its own column.

NOTE: You can work with data in a number of different programs. If you do not have access to statistical software, you can use Microsoft Excel to prepare your data. We have included some tips throughout this guide that will help you manage your data in Excel.

STEP 1. IDENTIFYING DATA SOURCES

You will use multiple types of data in your analysis. A checklist and description of the data you should have follow. At the end of this guide are examples showing what each data set should look like.

- **Treatment status (required).** This variable is in the data set that you downloaded from the Coach after your technology users and your comparison group were randomly selected. It indicates which participants received access to the educational technology and which did not. Typically, a 1 in this column indicates access to the technology (treatment or intervention group) and a 0 indicates no access to the technology (comparison group).

CAUTION: You must use the original file with the original treatment status variable, even if someone in your comparison group found a way to use the technology, or someone in your treatment group did not use it. The Coach will calculate the effect of being randomly assigned to use the technology on your outcome of interest, regardless of how much exposure participants actually received. If you move participants from one group to another after randomization, your groups might be unbalanced, and you will not be able to make reliable conclusions about the technology's effect.

- **Outcome data (required).** These are data on the outcome you are using to determine the effect of your educational technology, such as test scores.
- **Pre-test data (recommended).** This is a measure of the outcome before the introduction of the technology (such as an assessment from the beginning of the school year). Pre-test data are necessary to create similar groups if you are not using random assignment and these data can also be useful to control for pre-intervention abilities in a randomized pilot.
- **Background characteristics (recommended).** Background characteristics provide data on observable traits for each participant. These could include grade level, gender, ethnicity, individualized education program status, English learner (EL) status, socioeconomic status (SES), and more. It is necessary to include background characteristics in your analysis if a lot of participants drop out of one group. (For instance, if all students remain in the technology user group but 25 percent drop out of the comparison group, you might suspect that the two groups now differ). Accounting for background characteristics can rebalance the two groups. However, the Coach recommends that you include some background characteristics in your analysis even if your evaluation went as planned.

CAUTION: Some of these characteristics, such as EL status, can change over time. It is important to know if these characteristics were recorded before or after the introduction of the educational technology because some technologies might affect them (for example, the introduction of a reading software). *It is preferable to measure all background characteristics before the introduction of the educational technology.*

STEP 2. PROCESSING YOUR DATA

After you have identified the data elements and data sources, the second step is to combine all data elements into one **tidy data**[1] set and prepare the variables that you will use for analysis. (We explain below how to do this.) We recommend generating tidy data sets not only because it is a requirement to use the RCE Coach and most statistical software packages for analyses, but because a tidy data set is easy to manipulate, model, and visualiz The data set at the beginning of this guide is an example of a tidy data set. This section will take you through a series of questions that will help you create your own.

CONSIDER: It is important that all of your data are recorded using individual identifiers. These are unique codes for each participant. The identifiers are what will enable you to combine (merge) data sets. These could be a student ID number, school ID number, teacher ID number, and so on. If you have students in multiple classrooms or teachers in multiple schools, you should include an ID for both the individual (that is, student or teacher) and group (that is, class or school).

A. Is each observation a row and each variable a column?

Exhibit 2. Example: Columns and rows

AnonStudentID	Treatment	Fall score	Spring score	Gender
159508	1	320	35	F
694677	0	450	52	M
807588	0	999	37	F
482489	1	410	89	M

NO: Reorganize your data set so that each row represents one observation. Each variable you are interested in should be its own column.

NOTE: The data set that you downloaded from the Coach after random assignment might look different from the data set you now need for your analysis. The analysis data set should contain one row for each individual or group for which you measured outcomes.

For example, if you randomly assigned students to use the technology and you want to look at student test score outcomes, you probably have to add only the post-test scores to the original data set of students. However, if you randomly assigned teachers or classes to use the technology

[1] Having tidy data means that you've used a standardized way to structure your data set. Specifically: Each variable forms a column; each observation forms a row; each type of observational unit forms a table. To learn more about tidy data, see Wickham, H. "Tidy Data." Journal of Statistical Software, vol. 59, no. 10, 2014, pp. 1–23. doi:http://dx.doi.org/10.18637/jss.v059.i10

and you want to look at student test score outcomes, your original data set had one row per teacher, but this data set should have one row per student. You must include both student and teacher or class IDs. Also, student characteristics should be individual to each student, not the average of each class.

YES: Continue on to B.

B. Do you have one data set that contains all of the variables you will need for analysis?

NO: You will have to merge your existing data sets into one complete data set. This will be easy to do using the unique identifiers. If you are using Excel to manage your data you can do this using a VLOOKUP function.

CAUTION: Some observations might be present in some data sets but not in others. Therefore, when merging these data sets you could introduce some missing data. For example, a student in a data set consisting of test scores might not exist in another data set and therefore could have missing data for other variables (such as treatment status and background characteristics) after combining the data sets.

YES: Continue on to C.

NOTE: If you are using a statistical software package you will want to make sure that all of these variables are recognized as numeric values and not string or character values.

C. Are all of the categorical variables that your analysis will use numeric?

NO: Convert all of your categorical variables, or variables that include names or labels, into numbers. This might mean you have to change a variable into a binary or dummy variable. A dummy variable uses 1 to indicate yes or that a condition was met and 0 to indicate no or that a condition was not met.

For example, if your background characteristics include gender as a variable, you might have male or female, or M or F in each cell of that column. Instead, you should change the variable from Gender to Female, and change each cell that indicates the participant is a female to 1, and each cell that indicates the participant is male to 0. You can do this for every variable that is non-numeric.

Exhibit 3. Example: Numeric variables

Student ID	Gender
159508	F
694677	M
807588	F

Student ID	Female
159508	1
694677	0
807588	1

NOTE: If your categorical variable contains more than two options, such as Race (in which the options are Asian, Black, White, and Other) you will have to create a binary or dummy variable for each option. For example, Asian would be one column (with 0 representing non-Asians and 1 representing Asians), Black would be a second column, White would be a third column, and Other would be a fourth column (Exhibit X).

YES: Continue on to D.

D. Are all missing data coded as NA in your data set?

NO: If you have merged data sets and there are missing data, make sure you are consistently coding those as NA to ensure that the Coach can analyze your data. You want to be extra careful to make sure missing data have not been given a number designation, such as 0 or 999. These values will get incorporated into the analysis.

YES: Congratulations! You have a tidy data set!

STEP 3. CHECKING THE QUALITY OF YOUR DATA

After constructing your data file and converting your variables, the final step is to check the quality of your data. You can run the following checks to identify potential data issues that warrant additional investigation:

- **Check the minimum and maximum values of variables.** This check helps to identify extremely low or high values that are outliers in your distribution or that might signal a special missing code that has to be converted to a missing value. You might want to check with someone who is familiar with the data to confirm the value range makes sense.

NOTE: If you are working in Excel you can use MIN and MAX functions to easily find these values.

- **Consider the impact of missing data.** The dashboards, and some statistical software packages, will automatically drop observations that contain missing data. You should try to understand why data are missing and how excluding students, classrooms, or schools with incomplete data will affect your results.

NOTE: If you are working in Excel you can sort and filter your data to view missing values. To determine exactly how many values are missing for a single variable you can use the COUNTIF function; to determine how many observations have at least one missing value you can use a nested COUNTIF with OR function.

EXAMPLE DATA SETS

Exhibit 4. Data set 1: Test scores

AnonStudentID	Fall score	Spring score
159508	320	35
694677	450	52
807588	NA	37
482489	410	89
124226	604	67
232721	378	59
834305	NA	58
490514	380	NA

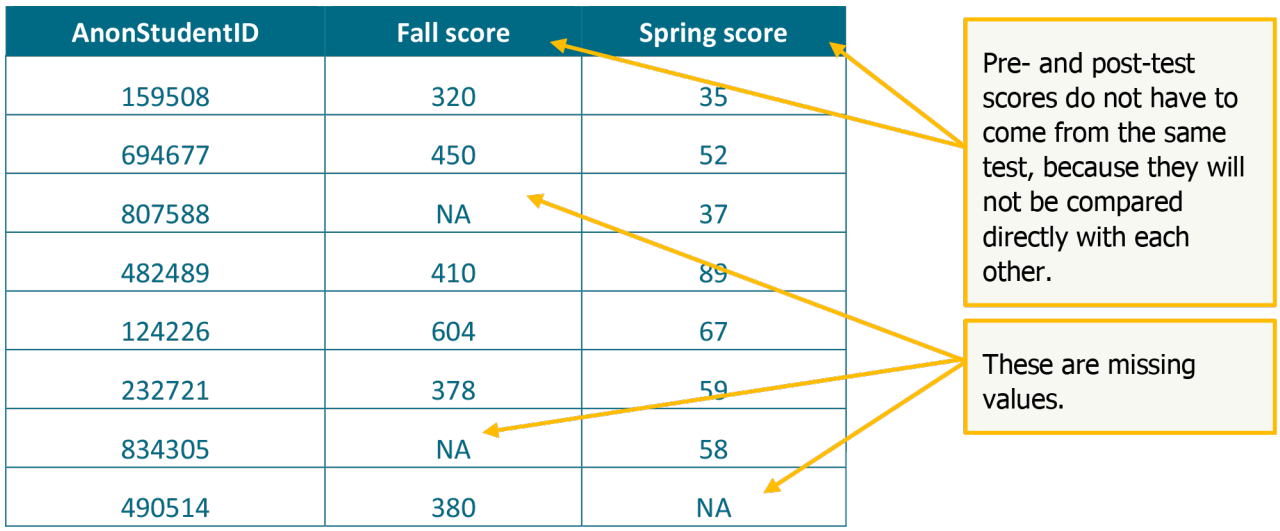


Exhibit 5. Data set 2: Treatment status

AnonStudentID	Treatment
159508	1
694677	0
807588	0
482489	1
555123	0
124226	1
232721	1
834305	1
490514	0
573401	0

Exhibit 6. Background characteristics (with non-numeric and numeric categorical variables)

AnonStudentID	SchoolID	Gender	EL status	SES	Race
159508	100	F	EL	Low	Black
694677	100	M	Not EL	Medium	White
807588	100	F	Not EL	High	Black
482489	100	M	EL	High	Asian
555123	100	F	EL	Low	Asian
124226	200	M	Not EL	Medium	Black
232721	200	F	Not EL	Low	Other
834305	200	F	EL	Missing	Missing

EL = English learner; SES = socioeconomic status.

Data set 3a: Background characteristics (with non-numeric categorical variables)

Data set 3b: Background characteristics (with numeric categorical variables)

AnonStudentID	SchoolID	Female	EL	Low_SES	Black	White	Asian	Other
159508	100	1	1	1	1	0	0	0
694677	100	0	0	0	0	1	0	0
807588	100	1	0	0	1	0	0	0
482489	100	0	1	0	0	0	1	0
555123	100	1	1	1	0	0	1	0
124226	200	0	0	0	1	0	0	0
232721	200	1	0	1	0	0	0	1
834305	200	1	1	NA	NA	NA	NA	NA

EL = English learner; SES = socioeconomic status.

The Coach requires that all categorical variables are converted to binary or dummy variables with values of 0 and 1. In the case of SES, you can choose to group medium and high SES together (1 = low and 0 = medium or high) if you are only interested in the effect of low SES on your outcome of interest.

NOTE: Research shows that students of higher SES often have an academic advantage over students of lower SES because of differences in early education access, home enrichment, levels of stress, food access, and many other factors. Therefore, if possible, it is good to control for SES when analyzing your data. However, some schools or districts might not have access to SES measures or might not be able to use them in this evaluation because of privacy concerns.

Guide to Using Matched Comparison Approach

Matching Overview

When random assignment is not possible, you can use matching to create a good comparison group to help you learn what works.

THE EVALUATION CHALLENGE

You want to test whether an educational technology is effective, but it is impossible to simultaneously observe what happens when an individual uses or doesn't use that technology. If you introduce a technology and watch what happens, you could notice improvements, for example, in student test scores. However, you *cannot* assume that the technology *caused* improved student outcomes. Many other factors (including regular classroom teaching, other programs, student maturation, and so on) could have contributed to the increases.

To overcome these challenges, it is important to compare a group of technology users to a group of nonusers, on the assumption that the only real difference between them is whether they are using the technology. However, comparing technology users and nonusers brings an additional set of challenges. When we make comparisons without trying to ensure similarities between groups, it is possible that those who use the technology differ in any number of ways from those who do not use the technology. For example, hard workers might be more likely to try a new technology, but they also perform better on tests. This might cause you to confuse the effect of the technology with the effect of working hard (because either could cause the technology users to outperform nonusers). Those differences can make an ineffective technology look effective, or vice versa.

MATCHED COMPARISON DESIGN

Solution. You can match educational technology users to similar nonusers using pre-test measures and background characteristics. After you have created two similar groups, you are comparing apples to apples—the only observed difference between users and nonusers is their exposure to the technology (though there might be unobserved differences). Then, if you see differences in outcomes (such as student achievement scores) you can be confident that the new technology is moving the needle.

How it works. Suppose you have a student, Jane, who uses the reading technology U-Read. Jane has a higher reading score on the 5th-grade state reading test than other students. To know if U-Read is having a positive effect on the reading score, you would like to be able to observe a Jane in a parallel universe. "Parallel Jane" is exactly the same but has no access to U-Read. If this parallel Jane scored lower on the 5th-grade state reading test, you could conclude that U-Read moved the needle for Jane.

Exhibit 1. Jane's reading scores

TREATMENT
JANE
5th-grade reading score: 430
4th-grade reading score: 410

With matching. Your initial comparison, without matching, compares Jane with the average student. Matching attempts to find something as close as possible to Parallel Jane. When successful, you can conclude that any differences in achievement are due to the technology, not other factors.

Exhibit 2. Potential comparisons

JOHN	JILL	JENNY	JODY
5th-grade reading score: 420	5th-grade reading score: 380	5th-grade reading score: 420	5th-grade reading score: 410
4th-grade reading score: 415	4th-grade reading score: 360	4th-grade reading score: 400	4th-grade reading score: 395

You can compare Jane with one of the four students shown in Exhibit 2. The better the match, the more confident you can be in your conclusion that U-Read is leading to higher test scores. *With whom should you compare Jane?*

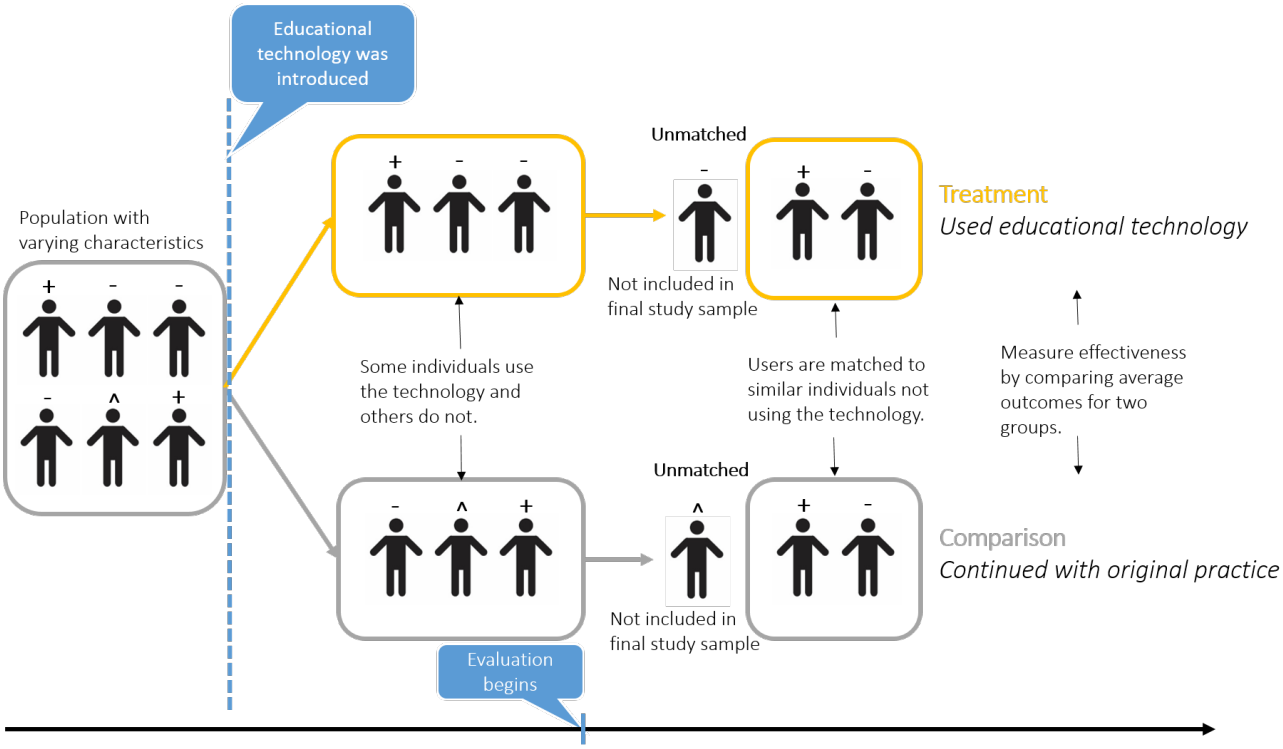
Assuming U-Read was introduced the first day of Jane's 5th-grade year, we want to match her with someone who had a similar 4th-grade reading score. Matching on a pre-test is fundamental for this technique to work. In this case, we would match Jane to John. Then we can use other statistical techniques to compare their 5th-grade achievement and determine if U-Read is moving the needle, or not, for Jane and others using U-Read.

In practice, you will want to use more than one observed characteristic to find matches for groups of students using a technology. For example, imagine that you observe that English is the second language for some of the students using U-Read and that others have individualized education programs. You might want to include these characteristics in your matching strategy because they are good predictors of who is using U-Read and of student achievement. To do so, the RCE Coach matching tool uses a statistical technique called nearest neighbor matching. When you create a valid comparison group using the RCE Coach, you can use the RCE Coach's impact estimation tool to determine whether U-Read is moving the needle.

Note: If the technology is targeted towards a very specific group (or if a very specific type of individual is likely to use the technology), the Coach will be less able to identify a good matched comparison.

For your RCE, the matching process will look something like the outcome in Exhibit 3.

Exhibit 3. The matching process



BASELINE EQUIVALENCE

An important question to ask is, "Was the matching successful?" A common way to assess this is to compare the average values of the groups' background characteristics after matching. For characteristics such as test score results from previous years or demographic characteristics, we can quantify the difference between the two groups using a measure called an effect size. Researchers use an effect size to measure different characteristics using the same yardstick. It is calculated by dividing the difference in means between the two groups by the standard deviation of the entire sample. The RCE Coach's matching dashboard automatically calculates the difference (measured as an effect size) between the users group average and the matched nonusers group average, for any variables you specify. For student outcomes, the U.S. Department of Education's What Works Clearinghouse (WWC) checks the equivalency of pre-intervention test scores according to the standards defined in Exhibit 4. Meeting the WWC standard for baseline equivalence helps bolster confidence that any effects you find are the result of the technology you studied.

Exhibit 4. WWC standards for baseline equivalence

Absolute value of difference between groups	WWC conclusion on baseline equivalence
effect size ≤ 0.05	Satisfies baseline equivalence requirement
0.05 < effect size ≤ 0.25	Requires statistical adjustment
0.25 < effect size	Does not satisfy baseline equivalence requirement

If the Coach's matching dashboard finds that your groups differ in an important baseline characteristic greater than 0.25 effect size units, it will prevent you from moving forward until you can develop a better a match.

CAUTION. The amount of confidence you have in the results of a matched comparison analysis is based on the similarity between the groups of users and matched nonusers. Two important notes follow from this:

1. Using observed characteristics to create a matched comparison will not necessarily yield two similar groups. It's important to think about the variables you include. Too few variables can lead to groups that aren't actually similar, and too many variables, particularly variables that aren't important for the outcome, will make it too difficult to match similar students. Focus on the variables that you think are important to the outcome or the likelihood of using the technology.
2. Even if you have a large set of observed characteristics, a matched comparison analysis cannot remove the possibility that individuals using the technology differ from those not using it in some unobserved way. For example, students using the technology might have parents who are more involved and advocate for additional attention or educational resources than students who are not using the technology. Or teachers who put in the extra effort to learn and use a new technology might work hard to enhance other aspects of their instruction as well. If so, the measured effect of the intervention could be inaccurate, even if the groups appear to be well matched on observed characteristics.

The only way to remove those potential differences between the two groups is to use random assignment to select the treatment and comparison groups. When properly conducted, random assignment ensures that the two groups are similar in both observed and unobserved characteristics.

Guide: Matching

When it's not possible to create treatment and comparison groups using random assignment, you can use a technique called matching that will create a group of nonusers that is similar to your group of users on observed characteristics.

DESCRIBE YOUR SAMPLE

Before you can match users to nonusers, you will have to make sure the Coach has all of the necessary information about who is or will be using the technology.

1. **Who is using the technology?** The answer should reflect who you identified as your users when writing your research question.
2. **Who is not using the technology?** The answer should reflect who you identified as your comparison group when writing your research question.
3. **Was the technology targeted in any way?** Sometimes technologies are implemented for a specific purpose, such as to provide additional support to low-performing students, English learners, or students on individualized education programs. If you targeted the technology to a specific type of user, you should answer yes to this question.

UPLOADING YOUR DATA

Before you upload your data, refer to our guide, Prepare Your Data for Analysis, to ensure that the file you upload meets the necessary requirements for the Coach to match your participants. Your data file should have one row for each of your participants and one column for each data field (for example, test scores and background characteristics).

The Coach suggests that you create an anonymous ID for each study participant in your file. Be sure to create a key or a file that links the original names or IDs to the new anonymous IDs. You will need this key to know who was assigned to each group and to add outcome data at the end of the pilot period.

Identifying variables

After you upload your data, you will have to tell the Coach which variable names correspond with key pieces of information.

- **Who is using the technology?** Choose the name of the data field that distinguishes who is using the technology from who is not (1 = users, 0 = nonusers).
- **Select your pre-test measure.** It is most important that users and nonusers had similar performance on your outcome of interest before they started using the technology. For

example, if you are interested in examining the effect of a technology on math achievement, you would want students in each group to have similar scores on a math test taken just before using the technology and you would use that math test as your pre-test matching variable and select which variable corresponds with scores here. If you are interested in a technology that improves students' attendance, you would want students in each group to have similar attendance records just before using the technology.

- **Which characteristics matter when matching users to similar nonusers?** Select characteristics of your students that could affect students' academic achievement. You can (and should) try to match your technology users to nonusers on more than one characteristic. For example, if you are matching students, you might want to match on individualized education program or English learner status, as well as characteristics of their teachers, such as years of experience, if the teacher plays an important role in students' use of the technology.

NOTE: You do not necessarily have to match students on every available characteristic. If you include too many characteristics, the Coach might not be able to find suitable matches. You should include only characteristics that you believe influence the outcome in which you are interested. For example, if your outcome of interest is student attendance, you might not have to include information about the students' teacher, such as years of experience, as it is unlikely to affect student attendance.

- **Should matched students always be in the same grade?** In some cases it is important that matched pairs of students are always in the same grade to ensure comparability. For example, if the pre-test that students take differs by grade, you would want to match by grade. If this is not relevant or is not a concern for your evaluation—for example, if you're using a vertically scaled assessment—leave match any grade selected. Otherwise, please select the name of the data field that indicates the student's grade.

Prepare Your Data for Analysis: Matched Comparison Design

INTRODUCTION

To answer your research question you will need data from multiple sources. If you need guidance on processing your data, this document will take you through best practices. At the end you will have a data file that will look like this and be ready to upload to the Coach:

Exhibit 1. Summary table

Missing data are replaced with NA so that the Coach doesn't include a value for that observation

Categorical variables must be converted to binary variables with two possible values: 1 instead of yes and 0 instead of no.

AnonStudentID	SchoolID	Grade	Treatment	Fall_ Score	Spring_ Score	Female	EL	Low_SES	Black	White	Asian	Other
159508	100	3	1	320	35	1	1	1	1	0	0	0
694677	100	3	0	450	52	0	0	0	0	1	0	0
807588	100	4	0	NA	37	1	0	0	1	0	0	0
482489	100	4	1	410	89	0	1	0	0	0	1	0
834305	200	4	1	NA	58	1	1	NA	NA	NA	NA	NA
490514	300	3	0	380	NA	NA	NA	NA	NA	NA	NA	NA
573401	300	3	0	468	45	NA	NA	NA	NA	NA	NA	NA
275321	300	4	1	523	78	NA	NA	NA	NA	NA	NA	NA
288475	300	3	0	375	37	NA	NA	NA	NA	NA	NA	NA

EL = English learner; SES = socioeconomic status.

Each row represents a single observation (student, teacher, school, and so on).

Each variable has its own column.

NOTE: You can work with data in a number of different programs. If you do not have access to statistical software, you can use Microsoft Excel to prepare your data. We have included some tips throughout this guide that will help you manage your data in Excel.

STEP 1. IDENTIFYING DATA SOURCES

You will use multiple types of data in your analysis. A checklist and description of the data you should have follow. At the end of this guide are examples showing what each data set should look like.

CONSIDER: It is important that all of your data should be recorded using **individual identifiers**. These are unique codes for each participant. The identifiers are what will enable you to combine (merge) data sets. These could be a student ID number, school ID number, teacher ID number, and so on.

- **Outcome data (required).** These are data on the outcome you are using to determine the effect of your educational technology, such as test scores.
- **Treatment status (required).** This is the data set that indicates which participants received access to the educational technology and which did not. Typically, a 1 in this column indicates access to the technology (treatment or intervention group) and a 0 indicates no access to the technology (comparison group).
- **Pre-test data (required).** This is a measure of the outcome before the introduction of the technology (such as an assessment from the beginning of the school year). Pre-test data are necessary to create similar groups. If you are measuring an outcome other than student achievement, you should include a variable that is highly correlated with your outcome of interest.
- **Background characteristics (recommended).** Background characteristics provide data on observable traits for each participant. These could include gender, ethnicity, individualized education program status, English learner (EL) status, and socioeconomic status (SES). Background traits are particularly important for matched comparison evaluations because you can use them to make sure your intervention and comparison groups are well matched (balanced).

CAUTION: Some of these characteristics, such as EL status, can change over time. It is important to know if these characteristics were recorded before or after the introduction of the educational technology because some technologies could affect them (for example, introducing a reading software). *It is preferable to measure all background characteristics before introducing the educational technology.*

STEP 2. PROCESSING YOUR DATA

After you have identified the data elements and data sources, the second step is to combine all data elements into one **tidy**[1] dataset and prepare the variables that will be used for analysis. (We explain below how to do this.) We recommend generating tidy data sets not only because it is a requirement

[1] Having **tidy data** means that you've used a standardized way to structure your data set. Specifically: 1.Each variable forms a column; 2. each observation forms a row; 3. each type of observational unit forms a table. To learn more about tidy data, see Wickham, H. "Tidy Data." Journal of Statistical Software, vol. 59, no. 10, 2014, pp.1–23. doi:http://dx.doi.org/10.18637/jss.v059.i10

to use the RCE Coach and most statistical software packages for analyses, but because a tidy data set is easy to manipulate, model, and visualize. The data set at the beginning of this guide is an example of a tidy data set. This section will take you through a series of questions that will help you create your own.

A. Is each observation a row and each variable a column?

Exhibit 2. Example: Columns and rows

AnonStudentID	Treatment	Fall score	Spring score	Gender
159508	1	320	35	F
694677	0	450	52	M
807588	0	999	37	F
482489	1	410	89	M

NO: Reorganize your data set so that each row represents one observation. Each variable you are interested in should be its own column.

YES: Continue on to B.

B. Do you have one data set that contains all of the variables you will need for analysis?

NO: You will have to merge your existing data sets into one complete data set. This will be easy to do using the unique identifiers. If you are using Excel to manage your data you can do this using a VLOOKUP function.

CAUTION: Some observations might be present in some data sets but not in others. Therefore, when merging these data sets you could introduce some missing data. For example, a student in a data set consisting of test scores might not exist in another data set and, therefore, could have missing data for other variables (such as treatment status and background characteristics) after combining the data sets.

YES: Continue on to C.

NOTE: If you are using a statistical software package, you will want to make sure that all of these variables are recognized as numeric values and not string or character values.

C. Are all of the categorical variables that will be used for your analysis numeric?

NO: Convert all of your categorical variables, or variables that include names or labels, into numbers. This might mean you have to change a variable into a binary or dummy variable. A dummy variable uses 1 to indicate yes or that a condition was met and 0 to indicate no or that a condition was not met.

For example, if your background characteristics include gender as a variable, you might have male or female, or M or F in each cell of that column. Instead, you should change the variable from Gender to Female, and change each cell that indicates the participant is a female to 1, and each cell that indicates the participant is male to 0. You can do this for every variable that is non-numeric.

Exhibit 3. Example: Numeric variables

Student ID	Gender
159508	F
694677	M
807588	F

Student ID	Female
159508	1
694677	0
807588	1

NOTE: If your categorical variable contains more than two options, such as Race (in which the options are Asian, Black, White, and Other) you will have to create a binary or dummy variable for each option. For example, Asian would be one column (with 0 representing non-Asians and 1 representing Asians), Black would be a second column, White would be a third column, and Other would be a fourth column (Exhibit 7).

YES: Continue on to D.

D. Is all missing data coded as NA in your dataset?

NO: If you have merged data sets and there are missing data, make sure you are consistently coding that as NA to ensure that the Coach can analyze your data. You want to be extra careful to make sure missing data have not been given a number designation, such as 0 or 999. These values will get incorporated into the analysis.

YES: Congratulations! You have a tidy data set!

STEP 3. CHECKING THE QUALITY OF YOUR DATA

After constructing your data file and converting your variables, the final step is to check the quality of your data. You can run the following checks to identify potential data issues that warrant additional investigation:

- **Check the minimum and maximum values of variables.** This check helps to identify extremely low or high values that are outliers in your distribution or that can signal a special missing code that must be converted to a missing value. You might want to check with someone who is familiar with the data to confirm the value range makes sense.

NOTE: If you are working in Excel you can use MIN and MAX functions to easily find these values.

- **Consider the impact of missing data.** The dashboards, and some statistical software packages, will automatically drop observations that contain missing data. You should try to understand why data are missing and how excluding students with incomplete data will affect your results.

NOTE: If you are working in Excel you can sort and filter your data to view missing values. To determine exactly how many values are missing for a single variable you can use the COUNTIF function; to determine how many observations have at least one missing value you can use a nested COUNTIF with OR function.

EXAMPLE DATA SETS

Exhibit 4. Data set 1: Test scores

AnonStudentID	Fall score	Spring score
159508	320	35
694677	450	52
807588	NA	37
482489	410	89
124226	604	67
232721	378	59
834305	NA	58
490514	380	NA

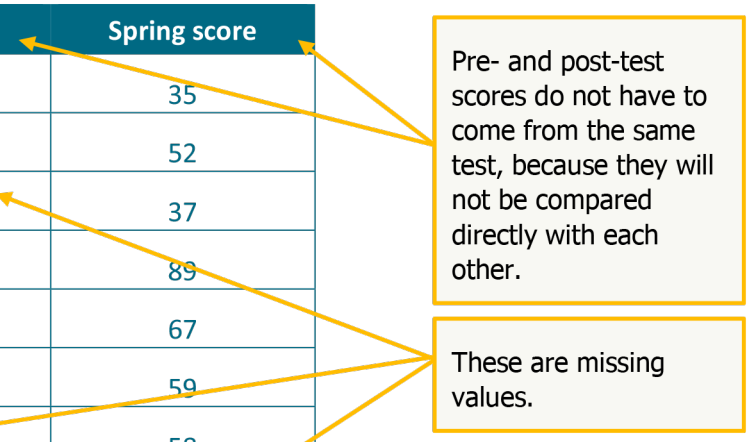


Exhibit 5. Dataset 2: Treatment status

AnonStudentID	Treatment
159508	1
694677	0
807588	0
482489	1

Exhibit 6. Background characteristics (with non-numeric categorical variables)

AnonStudentID	SchoolID	Gender	EL status	SES	Race
159508	100	F	EL	Low	Black
694677	100	M	Not EL	Medium	White
807588	100	F	Not EL	High	Black
482489	100	M	EL	High	Asian
555123	100	F	EL	Low	Asian
124226	200	M	Not EL	Medium	Black
232721	200	F	Not EL	Low	Other
834305	200	F	EL	Missing	Missing

Data set 3a: Background characteristics (with non-numeric categorical variables)

Data set 3b: Background characteristics (with numeric categorical variables)

EL = English learner; SES = socioeconomic status.

AnonStudentID	SchoolID	Female	EL	Low_SES	Black	White	Asian	Other
159508	100	1	1	1	1	0	0	0
694677	100	0	0	0	0	1	0	0
807588	100	1	0	0	1	0	0	0
482489	100	0	1	0	0	0	1	0
555123	100	1	1	1	0	0	1	0
124226	200	0	0	0	1	0	0	0
232721	200	1	0	1	0	0	0	1
834305	200	1	1	NA	NA	NA	NA	NA

EL = English learner; SES = socioeconomic status.

The Coach requires that all categorical variables are converted to binary or dummy variables with values of 0 and 1. In the case of SES, you can choose to group medium and high SES together (1 = low and 0 = medium or high) if you are only interested in the effect of low SES on your outcome of interest.

NOTE: Research shows that students of higher SES often have an academic advantage over students of lower SES because of differences in early education access, home enrichment, levels of stress, food access, and many other factors. Therefore, if possible, it is good to match students with similar levels of SES and to control for SES when analyzing your data. However, some schools or districts might not have access to SES measures or might not be able to use them in this evaluation because of privacy concerns.

Appendix A:
Usage Data Case Studies

Fictional Case Study 1

Lebanon Public Schools (LPS) recently acquired the reading comprehension system TigerRead. Administrators at LPS are interested in examining how teachers in the district are implementing TigerRead. Specifically, administrators wish to know whether teachers actually used the system with their students intensively enough that impacts on learning would be expected. The administrators form the question "How often is TigerRead used in classrooms throughout the fall semester?"

To answer this question, LPS administrators had to get specific usage data from the TigerRead developers. In thinking about the exact data needed to answer their question, LPS administrators realized that they weren't sure whether the total amount of time or the number of sessions working with the system was more important. They decided to ask the TigerRead developers for four pieces of information:

1. A classroom identifier
2. A metric of time
3. The total number of sessions associated with each time period
4. The average amount of time spent in the system during the corresponding time period

Each of these pieces of information played a crucial role in helping LPS administrators answer their question.

The **classroom identifier** helped them examine use across classrooms; thus, LPS administrators could see which classrooms used the learning system to different extents and could understand the variation in implementation patterns.

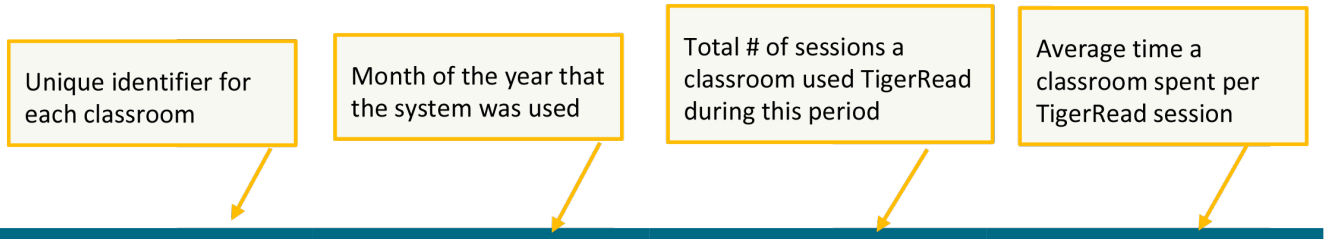
A **metric of time** (such as day, week, or month) provides the unit of analysis for examining intensity of system use. The smaller the unit, the more data administrators will have to analyze, but if units are too coarse (say semester or year), they could miss important differences in usage patterns.

Total number of sessions per time period enables the administrators to examine the frequency of use. This can be useful for examining how TigerRead is used across time. (That is, do teachers use the system mostly at the beginning or end of the semester? Is it used in many short sessions or a few long sessions?)

Finally, the **average time spent** in the system during the corresponding time period provides a view of the length of time that students worked on the system during that period.

In our example, administrators at LPS worked with developers at TigerRead to identify the specific variables and data fields needed to answer their question (see Exhibit 9).

Exhibit 9. Specific usage data



Classroom ID	Month of year	Total # of sessions	Average time per session
123	September	10	25
123	October	8	30
123	November	6	45
123	December	0	0
456	September	0	0
456	October	5	34
456	November	8	20
456	December	10	30
789	September	7	20
789	October	7	25
789	November	7	24
789	December	7	27
467	September	2	10
467	October	10	20
467	November	3	30
467	December	10	35

When LPS administrators had these data, they had to think about the best ways to analyze the data to address their question. They started, as many data analyses do, by compiling **descriptive statistics**.[1] These included classroom averages, medians, and ranges for the number and length of session for each of the four months in the data set.

In the previous example, LPS administrators wanted to examine how often TigerRead was used within district classrooms. A good introductory analysis that LPS administrators can conduct to investigate their question would be to examine the overall average number of minutes that each class used the

[1] **Descriptive statistics**. Statistics that describe, show, or summarize data in a meaningful way

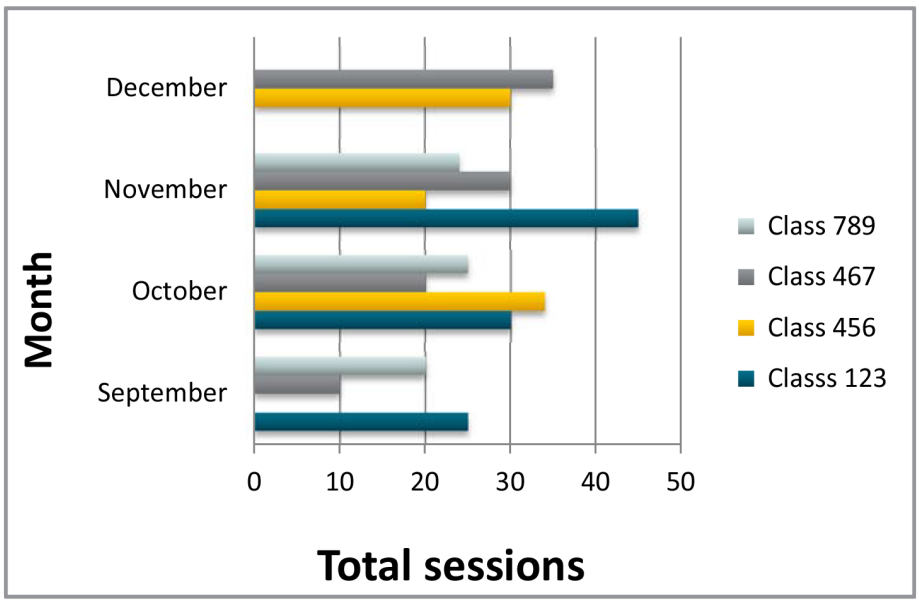
system each month. Using the descriptive statistics on (1) average time spent per class, (2) average time spent per month, and (3) their corresponding ranges, LPS administrators generated a descriptive data report (Exhibit 10) and visual charts (Exhibit 11). Administrators can use these reports to better understand *how* and *when* teachers used TigerRead.

Exhibit 10. Summary of TigerRead use (sessions per month)

Class ID	September	October	November	December	Average across months	Range across months
123	25	30	45	0	25	0–45
456	0	34	20	30	21	0–34
467	10	20	30	35	23.75	10–35
789	20	25	24	27	24	20–27
Monthly averages	**13.75**	**27.25**	**29.75**	**23**	**23.44**	

Exhibit 11. Total sessions per month

Using Exhibits 10 and 11, LPS administrators can get a sense of the use of TigerRead. For instance, using Exhibit 10, administrators can see that overall TigerRead was used an average for 23 sessions per month across all four classes. They can also use Exhibit 10 to see the monthly breakdown of average sessions in the class per system. These data provide insight into how each class is using TigerRead across time. For example, it might be useful to learn that classroom 789 had the highest



average usage but the narrowest range of use, suggesting more consistent adoption than class 123, which had a high upper bound but a large range and relatively low average.

Similarly, the bar graph presented in Exhibit 11 provides a visual chart of the descriptive statistics. Through this exhibit, administrators can visually compare each class's use of TigerRead to the others across months in the fall semester. The exhibit shows that patterns of use varied substantially across classes.

Fictional case study 2

LPS administrators are interested in examining students' learning growth within the TigerRead system. Specifically, administrators wish to know how students' reading comprehension skills change over time and how they can track those changes within the system. LPS administrators develop the question *"How does students' use of TigerRead correspond to changes in their reading comprehension ability?"*

Similar to Example 1, LPS administrators need to get specific usage data from the TigerRead developers to answer this question. Again, it is important to think about the exact data that you need to answer your question. For the current example, LPS administrators need at least four pieces of information:

5. A student identifier
6. A metric of time
7. Activity type
8. Metrics of performance within the system related to reading comprehension

Each of these pieces of information plays a crucial role in helping LPS administrators answer their question.

The **student identifier** allows for the examination of each student's specific use and performance within TigerRead. Thus, LPS administrators could see which students were using the learning system to different extents and could understand how usage patterns varied.

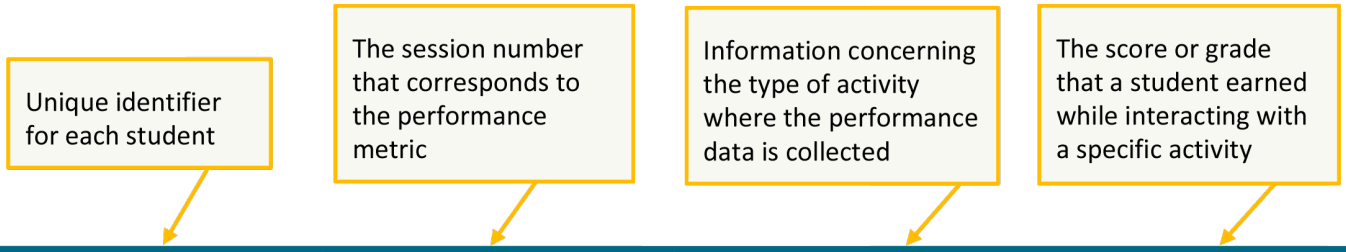
A **metric of time** (such as day, week, or month, can also be number of sessions) indicates how often (and at what frequency) each student used TigerRead.

Activity type provides metadata concerning the type of activity a student engages in (where the performance data are collected).

Metrics of performance within the system examine performance and thus growth across time. This can be useful for examining changes in reading comprehension as students use the TigerRead system.

In our example, administrators at LPS worked with developers at TigerRead to identify the exact variables and data fields needed to answer their question (Exhibit 12).

Exhibit 12. Variables and data fields

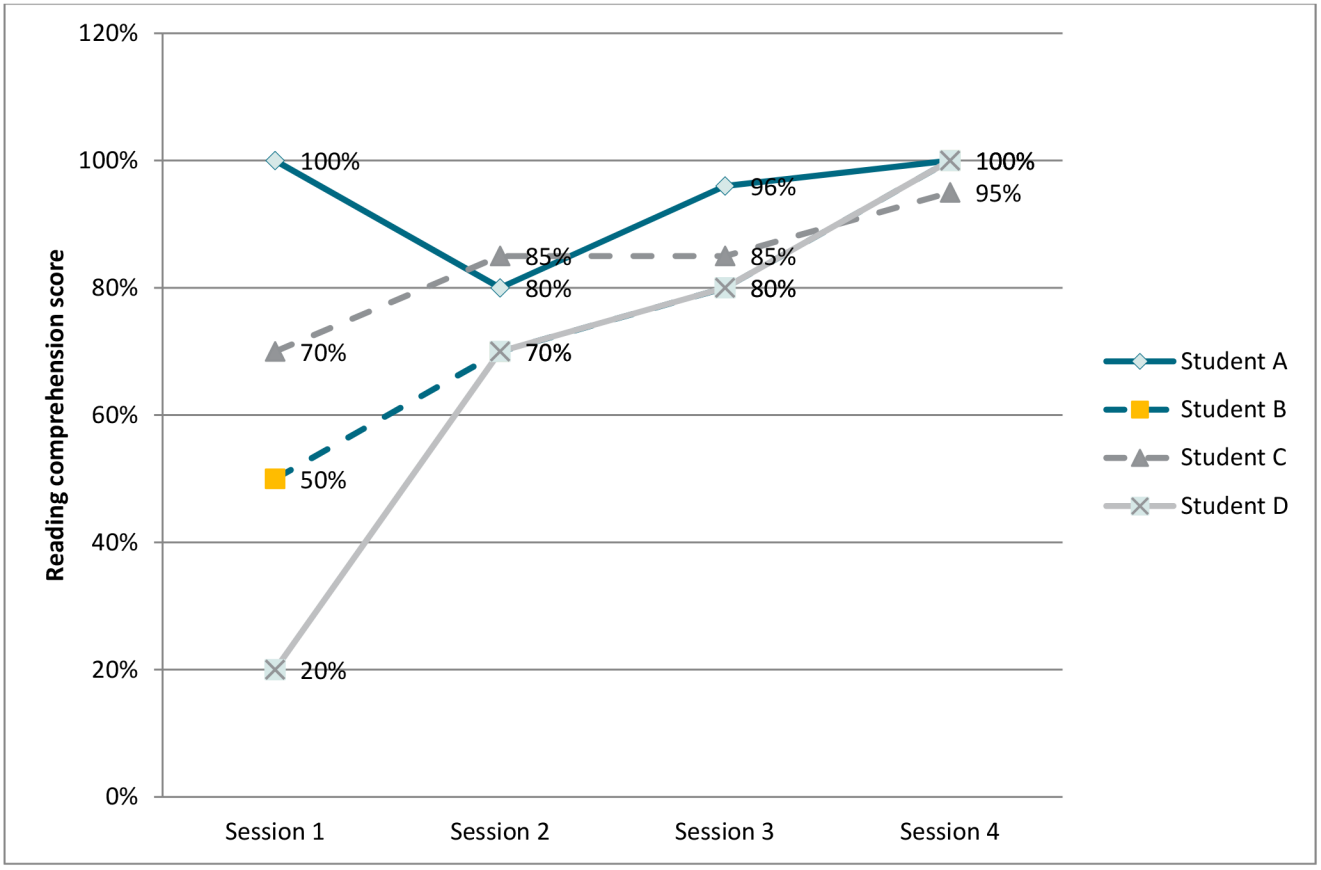


Student ID	Session number	Activity type	Performance metric
Student A	1	Quiz	100%
Student A	2	Checkpoint	80%
Student A	3	Quiz	96%
Students A	4	Quiz	100%
Student B	1	Quiz	50%
Student B	2	Checkpoint	70%
Student B	3	Quiz	80%
Student B	4	Quiz	100%
Students C	1	Quiz	70%
Students C	2	Checkpoint	85%
Students C	3	Quiz	85%
Students C	4	Quiz	95%
Students D	1	Quiz	20%
Students D	2	Checkpoint	70%
Students D	3	Quiz	80%
Students D	4	Quiz	100%

Again, after data are obtained, LPS administrators have to think about the best ways to analyze the data as a means to answer their specific question.

In this example, LPS administrators are interested in growth over time. Specifically, they want to examine how students' reading comprehension skills change over time and how to track those changes within TigerRead. An introductory analysis that LPS administrators can conduct to investigate their question would be examining trends (both growth and decreases) in performance across time by plotting each student's performance by session. Using the variables (1) students ID, (2) session number, and (3) performance metric, LPS administrators can generate a visual chart (see Exhibit 13) to better understand changes in students' performance across time.

Exhibit 13. Linear changes in reading comprehension performance



Using Exhibit 13, LPS administrators can get a sense of how students' performance in TigerRead changed across time. For instance, administrators can see that, overall, students' performance seemed to improve over time, with each student's end performance (session 4) being higher than his or her starting performance (session 1). Moreover, the data show that students initially had wide gaps in performance within the system, but that these gaps nearly closed after four sessions.

REFLECTION

In the previous examples, LPS administrators were able to work with TigerRead developers to successfully extract usage data to answer their questions. However, this process is not always as smooth as demonstrated here. Indeed, there is often a need to iterate and refine the data that are pulled from the system. The need for this kind of iteration often stems from difficulty communicating exactly what data you're referring to when the person you're talking to has vastly different experiences with the learning system software and with educational decision making than you do.

One way to avoid this extra work (and frustration) is to find common terminology with developers and to express the ultimate goal of using the data. For instance, asking a developer for time-stamped data will not provide you with the specific information you need to examine the question for example 1: "How often is TigerRead used in classrooms throughout the fall semester?" Instead, it is important to express the exact nature of your question and the variables and fields you wish to examine.

SOME COMMON DATA TERMS

This short glossary is designed to clarify the various types of data terminologies you could encounter and associated definitions. Please note that there is overlap between some terminologies, as many are used interchangeably.

GLOSSARY

Clickstream data Information related to the sequential order of users' events and mouse clicks in a system

Event Any discrete action a user takes within a system, such as submitting responses to an assessment or viewing an instructional video

Keystrokes Information related the keys struck on a keyboard during users' time in a system

Learning data Describes users' interactions with a system indicating student achievement on formative and summative assessments and/or capturing students' learning-related behaviors

Metadata points Information related to the content or topic addressed by a particular unit within a program

Natural language processing (NLP) Information related to users' discourse with and in a system (that is, written or oral discourse)

Process data Information related to a series of actions or events in a system

Session The time a user spends interacting with a system during one sitting, often between logging in and logging out

System or log data Information concerning the actions and events that users engage in when they interact with the system

Time stamps Information about when an event took place (for example, log-ins, log-outs, completed actions; often includes dates and times

Usage data Describes all types of user interactions with a system

Mouse tracking Information related the use and movement of a user's computer mouse while in a system

Telemetry Information related to how a system is used; often associated with technical aspects of the system (for example, browser, errors, run time, and number of users on the system)

Appendix A. Template A: Summary data

Student ID	Class or teacher ID	School ID	Grade	Total # of sessions	Total # of minutes	Average minutes per session	Total # of assessment attempts	Average score on assessments	Minimum score	Maximum score

Appendix B. Template B: Usage data, by topic area

Student, class, or school ID	Unit, topic, module, or lesson name	% of students attempted	% of students completed	Average score on assessments	Average minutes of use	Minimum score	Maximum score

Appendix C. Template C: Usage data, by time period

Student, class, or school ID	Period 1 # of sessions	Period 1 total # of minutes	Period 2 # of sessions	Period 2 total # of minutes	Period 3 # of sessions	Period 3 total # of minutes	Period 4 # of sessions	Period 4 total # of minutes

Appendix B:
Technical Appendixes

Random Assignment Technical Appendix

This document provides a technical overview of how the RCE Coach Random Assignment tool functions. For a brief introduction to randomized pilots as a research design, refer to our Random Assignment Overview.

Random assignment attempts to create two groups that are as similar as possible in both observed and unobserved characteristics. You can think of the random assignment process as flipping a coin to assign students to either pilot the technology or not. Using a fair coin, each student would have a 50 percent chance of being assigned to the group piloting the technology. Therefore, if you have a sufficiently large pool of potential users, the two groups should have the same observed and unobserved characteristics. You can think of the Coach as a way to flip a virtual coin many times, very quickly. In addition, anyone can access the code to see that the coin is truly random, and the chance of being assigned to the treatment group can differ from 50 percent to provide a smaller, or larger, fraction of the population to initially test the technology.

WHAT DOES IT MEAN TO HAVE A SUCCESSFUL RANDOM ASSIGNMENT?

A successful random assignment process will produce treatment and comparison groups that are very similar. When the groups are large, random assignment is likely to succeed at creating two similar groups. For smaller samples, it is possible that the two groups differ in important ways even though a random process was used to create them. The Coach uses the baseline equivalence standard from the What Works Clearinghouse to assess whether the two groups are similar enough. We encourage you to use some pre-intervention measure that is related to the outcome to check for baseline balance, but this is not required. The Coach will consider that random assignment worked as expected if the differences on all the variables selected to control for balance are less than 0.25 standard deviations; otherwise, the Coach will attempt the random assignment process again. The Coach will assume that random assignment worked as expected if no variables are specified to check for baseline equivalence.

HOW DOES THE TOOL WORK?

The RCE Coach uses R, a free software package for statistical computing. In particular, the Coach uses the base R function sample.int to perform the random assignment. The Coach will first perform several checks on the data to confirm that there are no data issues that would cause an error with the random assignment code. Then the Coach will randomly assign observations to the intervention or comparison group and if baseline variables were specified, will check that random assignment worked as expected. The rest of this document describes the specific data checks and random assignment algorithm.

Ed Tech Rapid Cycle Evaluation Coach

Step 1: Check for data issues

The Coach will perform the following checks to verify that the Coach has the necessary inputs to perform random assignment:

1. Data are NULL, not readable, or have 0 observations.
2. Variable specifying the randomization unit ID is not specified or does not exist.
3. Whether groups should be determined by either a set number or percentage is not indicated.
4. Percentage was selected for assignment and the number of users indicated is less than or equal to 0 or greater than or equal to 100.
5. Number was selected for assignment and the number of users indicated is less than or equal to 0 or greater than or equal to the number of records in the data.
6. Baseline variables are specified and not all exist in the data.
7. Baseline variables are specified and one or more are not numeric.
8. A variable named Treatment already exists in the data.

If any test fails, the Coach will print a message to help the user identify the problem. If the Coach detects no problems, it will proceed to Step 2.

Step 2: Randomize and check for balance

The Coach will perform the random assignment using a loop that stops either when random assignment is successful or after 10 failed attempts. Random assignment is deemed successful when all variables selected for the baseline equivalence checks are balanced across the two groups. If the process was successful, the user will be able to download the data and move to the next step in the Coach. Otherwise, the coach will randomize again with a new random assignment seed. If the Coach fails to produce a successful random assignment 10 times, it will inform the user that the process did not work as expected. The following steps describe the loop:

1. The Coach randomly selects a random assignment seed.
2. Using sample.int a number of observations are chosen to be in the treatment group. This is either a fixed proportion of the sample (rounded up to the nearest whole observation) or a fixed number of observations.
3. The Coach checks if random assignment was successful.
4. If the process is successful a file will be available to download with the original data plus one variable indicating whether the observation belongs to the treatment or comparison group.

5. If the process is not successful
    a. If the loop has run fewer than 10 times, a new seed will be chosen and the loop will run again.
    b. If the loop has run 10 times, a message will explain to the user that the random assignment was not successful.

The code for the RCE Coach is open source under the General Public License Version 3 and will be available soon on our github repository.

Matching Technical Appendix

This document provides a technical overview of how the RCE Coach matching tool functions. For a brief introduction to matching as a research design, you can refer to our Matching Overview.

Matching attempts to group students (or the individuals you are trying to affect) with similar characteristics. The ideal match would be two students identical on every observed and unobserved characteristic. In practice, it is very unlikely that two students will be identical on all observed characteristics, so you must determine which observable characteristics you will match on based on how likely those characteristics are to affect the outcome of interest.

Matching on only one variable is conceptually simple. Suppose that you have a set of students using U-Read, an educational technology, and another set not using it. You are trying to create a matched data set using only a pre-test score. A simple algorithm would look like this:

1. Define a distance measure. For example, the absolute distance between pre-test scores of 240 and 265 is: |240-265|= 25
2. Take a student using U-Read and calculate the distance between his pre-test and the pre-test of each of the students not using U-Read.
3. Match this student to the student with the minimum distance.
4. Repeat Steps 2 through 4 for all students using U-Read.

It is more complicated to define distance when you are trying to match students using more than one characteristic. When you use multiple matching variables, you have to summarize this information into a single distance number. For example, suppose that you want to include an indicator for English as a second language and another for economic disadvantage. In this case, we can create a measure that summarizes all this information and match based on the summary. We can combine all of this information, even though the scales differ, by calculating the probability that a student with a given set of observable characteristics uses the educational technology. The goal is that after matching, students in the intervention group will be matched to students from the comparison group who look the same on observable key characteristics.

HOW DOES THE TOOL WORK?

The RCE Coach uses R, a free software package for statistical computing. The Coach will first perform several checks on the data to confirm that there are no data issues that would cause an error with the matching code. Then the Coach will create a matched data set using the R package MatchIt. The rest of this document describes the specific data checks and randomization algorithm.

Ed Tech Rapid Cycle Evaluation Coach

Step 1: Check for data issues

The Coach will perform the following checks to verify that the Coach has the necessary inputs to randomly assign the students:

1. Data is NULL, not readable, or has 0 observations.
2. No treatment variable specified.
3. No matching variables specified.
4. One or more matching variables not found in the data.
5. Treatment variable has values other than 0, 1, and NA.
6. One or more matching variables are not numeric.

If any test fails, the Coach will print a message to help the user identify the problem. If the Coach detects no problems, it will proceed to Step 2.

Step 2: Create a matched data set

To create a matched data set, the RCE Coach uses one-to-one nearest neighbor matching (Rubin 1973). The Coach will select a comparison individual with the smallest distance for each individual using the educational technology.

After the Coach has selected matches for each individual using the educational technology, it will assess whether the two groups—the intervention group and the matched comparison group—are similar on these key characteristics. The Coach uses the baseline equivalence standard from the What Works Clearinghouse to make this assessment. We consider the matched group to be a valid comparison group if the baseline differences on all matching variables are less than 0.25 standard deviations.

If the Coach fails to generate a valid comparison group, it will define a maximum distance, known as a caliper, and repeat the matching algorithm. If the distance between the individual using the technology and the nearest individual not using it is larger than this caliper, the Coach will drop this observation. If the Coach cannot generate a valid comparison group with this caliper, it will use a smaller one. The Coach will start by using no caliper, and then move from 1.00 standard deviation of the distance measure to 0.25 standard deviations in 0.25 intervals. Using a caliper can mean that some users of the educational technology will not receive a match. If this is the case, you should be cautious when interpreting the results because you are measuring the effect only for the individuals who have a good match. This algorithm can be summarized as follows:

1. The Coach attempts to match treatment and comparison observations using nearest neighbor matching.

2. If matching in Step 1 successfully produced balanced samples, the matched groups are returned as a downloadable file and relevant sample sizes and graphic displays of balance are shown on the screen.
3. If matching in Step 1 was not successful, the Coach attempts to match using caliper matching, with an initial caliper value of 1.00 and decreasing by 0.25 until either matching produces balanced samples or the caliper reaches 0.
    a. If matching was successful, a downloadable file is available as in Step 2 and the same sample sizes and graphics are shown on screen.
    b. If matching was not successful for any caliper value, balance graphics are shown but there is no file to download.

The code for the RCE Coach is open source under the General Public License Version 3 and will be available soon on our github repository.

Which variables should I include for matching?

You should always include some baseline, or pre-intervention, measurement that is related to your outcome of interest. In the case of U-Read, you would include pre-test scores as one of your matching variables. However, there are cases in which you might want to use additional characteristics for matching to improve precision.

In our example, if the number of students using and not using U-Read is very different (one group is substantially larger than the other), the literature recommends including:

- Variables that affect the likelihood of using U-Read
- Variables that affect the outcome you want to study (for example, achievement), controlling for use of the educational technology
  (Rubin and Thomas 1996; Heckman et al. 1998; Ho et al. 2007)

If you do not have substantially more students not using U-Read than the number of students using U-Read, you **should not** include all available characteristics (Ho et. al. 2007). The intuition is that as you increase the number of characteristics used for matching it might become too difficult to find a good match. For example, think about searching for a new apartment: if you include too many characteristics (location, bedrooms, laundry, pets, size, and so on), your search could end up too narrow and you won't find any results.

Should matching variables also be included as covariates in the analysis?

YES! Matching is seldom perfect—some imbalance often remains on at least some variables. Regression adjustment for the matching variables can mitigate any remaining imbalance.

REFERENCES

Heckman, J., H. Ichimura, J. Smith, and P. Todd. "Characterizing Selection Bias Using Experimental Data." No. w6699. Cambridge, MA: National Bureau of Economic Research, 1998.

Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart. "Matchit: Nonparametric Preprocessing for Parametric Causal Inference." Journal of Statistical Software, [vol. 42, no. 8], 2007. Available at http://gking.harvard.edu/matchit.

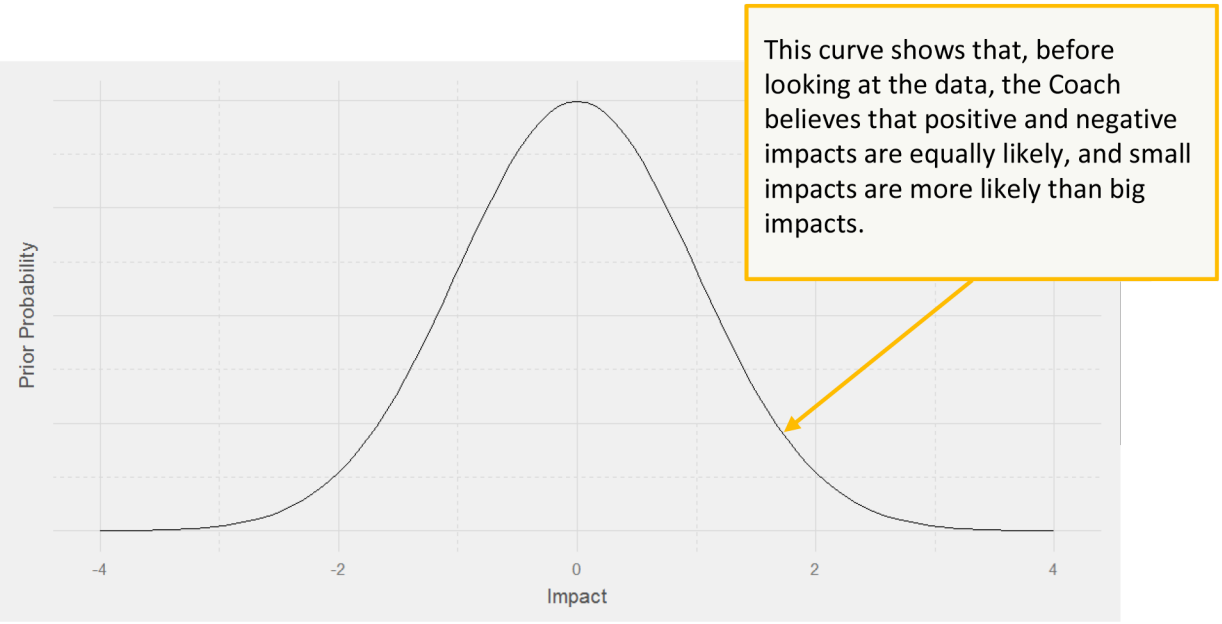
Rubin, D.B., and N. Thomas. "Matching Using Estimated Propensity Scores: Relating Theory to Practice." Biometrics, 1996, pp. 249–264. 7

Impact Estimation Technical Appendix

OVERVIEW

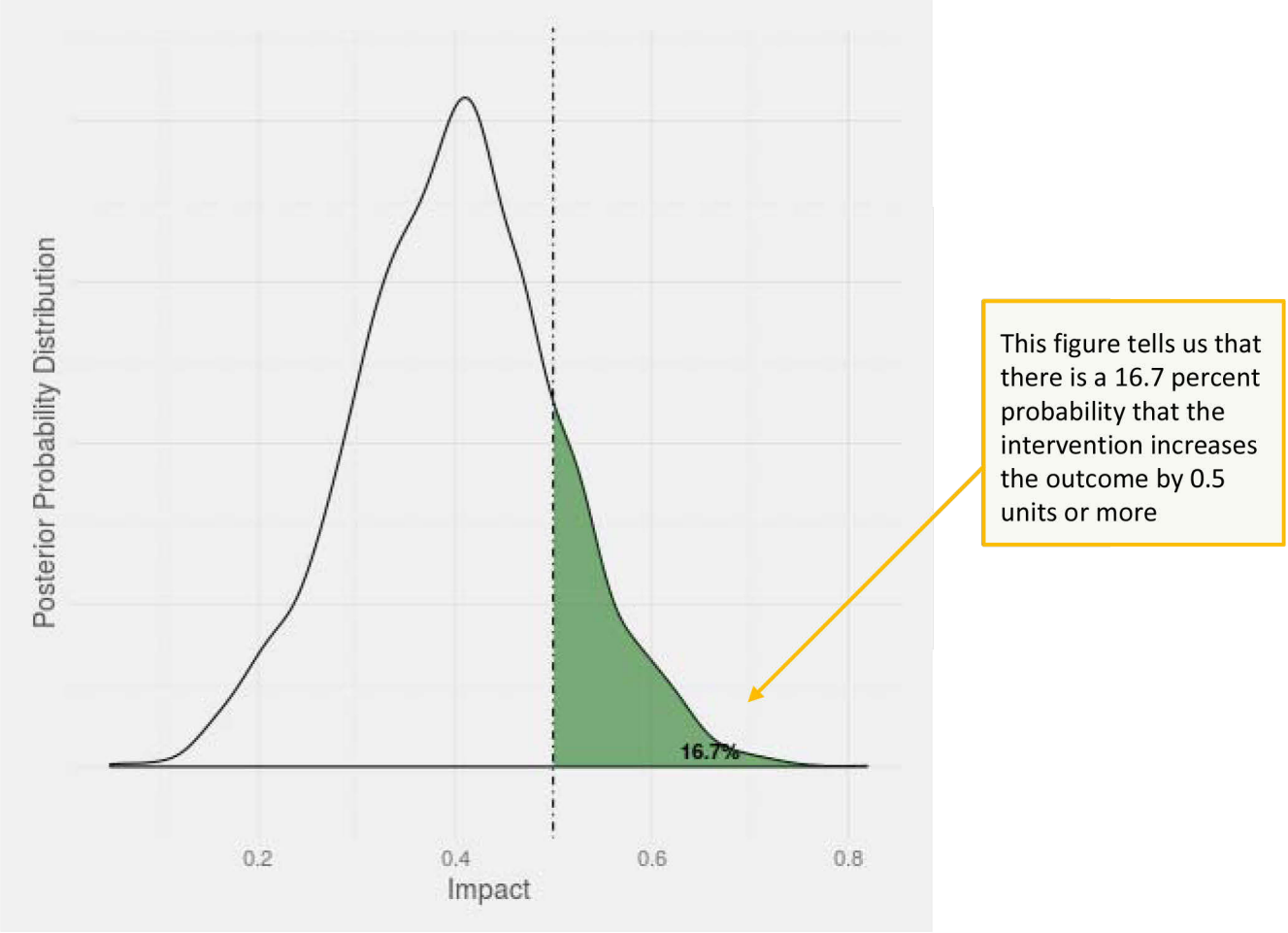
The RCE Coach will help you determine if an educational technology is moving the needle. Before you load your data, the Coach believes that negative and positive effects are equally likely, and that small impacts are more likely than big impacts. This is what statisticians call the prior distribution and it is depicted in Exhibit 1. A prior distribution describes the analyst's beliefs about a measurable outcome before any data are taken into account.

Exhibit 1. Prior distribution of the effect of the educational technology on the outcome



When the data are loaded and the other inputs are specified, the RCE Coach calculates the probability that the effect of an educational technology is above or below a threshold selected by the user (in Exhibit 2, the threshold, as indicated by the vertical line, is 0.5). This is the probability that the educational technology is moving the needle. Exhibit 2 illustrates this.

Exhibit 2. Probability that the educational technology increases outcomes by 0.5 units or more, given the data uploaded



The Findings Brief for each evaluation includes this information.

In this example, the user told the RCE Coach that any increase on the outcome larger than 0.5 units would be considered a success. The Coach calculated that there is a 16.7 percent probability that the educational technology moved the needle by at least that amount.

Note that the answer to whether the educational technology is moving the needle will depend on the threshold chosen by the user (0.5 units in this example) and how much uncertainty he or she is willing to tolerate. For example, some users might conclude that a technology is moving the needle if there is at least a 75 percent probability that the technology meets the threshold they chose, whereas others might want to be more confident, leading them to select a higher probability. In the previous example, if the user sets a certainty level of 95 percent, then this technology would not be found to be moving the needle, because 95 is larger than 16.7.

THE RCE COACH

The RCE Coach uses a program called Stan (see Carpenter et al., 2016) to estimate a linear model using Bayesian statistics. In particular, the Coach uses the R package RStan. The Coach acts as a simple graphic interface to help the user construct a formula. The user has to select the outcome, treatment indicator, and any additional covariates that the user might want to include. Before estimating the effect of the intervention, the Coach temporarily standardizes the data using the R function "scale." Then, the Coach estimates the following model:

$$y_i = \alpha + \eta T + \beta x_i + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma)$$
$$\sigma \sim N(0,1)$$
$$\beta_k \sim N(0,1)$$
$$\eta \sim N(0,1),$$

where y_i is the outcome of interest for individual i, T is the treatment indicator, and x_i is a matrix with the other covariates.[1] The term η tells you the effect of the educational technology—that is, how much higher the outcome was for the average student using the educational technology, $T = 1$, compared with the average student not using it, $T = 0$. The Coach computes the posterior distribution of η, which describes our understanding of the educational technology's effect after observing the data. Using this distribution, the Coach can calculate the probability that the treatment effect is above or below a threshold selected by the user, as shown in Exhibit 2.

WHICH VARIABLES SHOULD I INCLUDE AS CONTROL VARIABLES?

You should include all the variables, measured before you implement the technology, that you think can affect the outcome of interest and for which you have good data. In the case of student achievement, it's common to include previous achievement and other characteristics, such as indicators for English as a second language, socioeconomic status, and so on.

In addition, sometimes you might have to include an element to the model to account for clusters in the data, such as classrooms or schools. This applies when students are assigned to use or not use the technology as groups, rather than individually. To correctly estimate the effect of the intervention, you will have to tell the Coach that you used clusters during Step 5. By giving the Coach this information, you enable it to account for the possibility that students belonging to a cluster can do better or worse just because they belong to that cluster and not because they are using the technology. For example, when entire classrooms are assigned to conditions, the quality of teachers

[1] In the clustered case, we add cluster-specific random effects to the regression model that's given.

in each classroom can affect outcomes, in addition to any effect of the technology. In this case, you would want to indicate the variable that identifies classroom clusters so that the Coach can account for classroom-based differences in outcomes.

The code for the RCE Coach is open sourced under the General Public License Version 3 license and will be available soon on our GitHub repository.

CITATIONS

Carpenter, Bob, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A. Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. “Stan: A Probabilistic Programming Language.” Journal of Statistical Software, in press.

Stan Development Team. “RStan: the R Interface to Stan, Version 2.10.1.” Available at http://mc-stan.org.

Appendix C:
Glossary and FAQ

Glossary

Attrition One or more participants or groups of participants drops out of the study sample. Attrition is particularly important for randomized pilot because a high dropout rate can mean that the two groups that were equivalent or balanced when they were randomly assigned are no longer balanced. It is important to consider both overall attrition (the dropout rate of the full sample) and differential attrition (the difference in the dropout rate between the two groups).

Background Characteristics Information on study participants—such as gender, English learner status, or years of teaching experience—that might be related to accessing the technology or with the outcome you are measuring. These characteristics are recorded as variables in your data set and are used to create comparison groups that are similar to each other.

Baseline The beginning of the evaluation, before the introduction of the intervention (that is, just before the technology is used).

Baseline Equivalence A demonstration of the similarity between the educational technology users and nonusers before you implement the technology. You achieve equivalence by comparing the average of each group on one or more observed characteristics.

Binary Variable See Dummy Variable

Categorical Variable A variable that has a limited number of specific values. For example, high, medium, and low test scores can be represented as 3, 2, and 1, respectively.

Causality The relationship between a cause and an effect. The Coach is designed to help you use research designs that enable you to assess whether the technology you are studying causes the outcomes you wish to achieve.

Cluster A group of individuals. For example, students are clustered within classrooms and within schools, and classrooms are clustered within schools. It is important to account for clusters in analyses when you are assigning technology based on groups because individuals within a group share experiences. For instance, all students within a class have the same teacher, and the quality of that teacher's instruction will affect students' outcomes in a way that differs from a group of students in a different class.

Comparison Group A set of participants who are similar to the technology users but who do not have access to the educational technology. In a randomized controlled trial, the comparison group is the set of participants who were not selected to pilot the technology. In a matched comparison group design, the comparison group is usually the set of participants who do not use the technology.

Confounding Factors A hidden factor that influences an outcome, when you think the cause is actually something else (such as the technology).

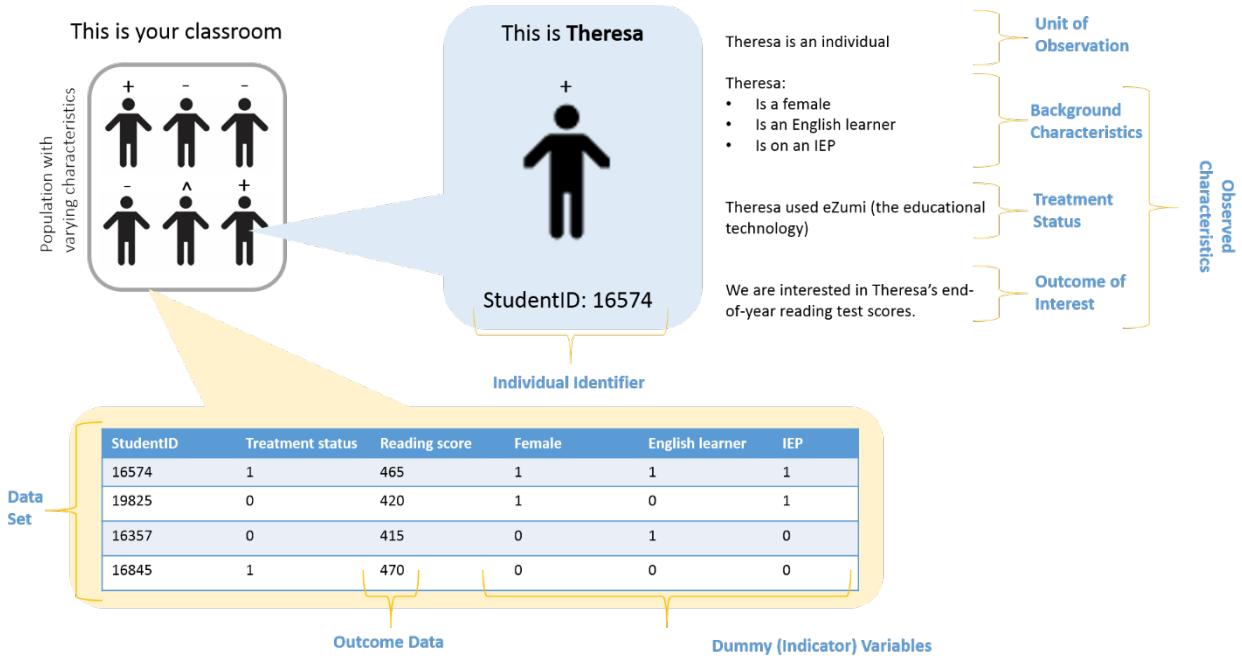
Control Variables See Covariates

Covariates Factors or characteristics that differ across participants and can be related to the outcome of interest. It is important to account (or control) for these characteristics in your analysis of a technology's effectiveness because one of these factors or characteristics—rather than the technology—could explain your findings.

CSV File A CSV (comma-separated values) file stores data in tables as plain text and is widely used to exchange or transfer information. The Coach uses the CSV format to read in your data files for analysis. You can save Microsoft Excel files as CSV files using the "Save As" function.

Data Set A collection of related information that is organized in a table of rows and columns. Each column represents a variable (for example, whether a participant used a technology, student test scores, and gender). Each row represents a different observation (for example, a student, a teacher, or a school). The Coach uses your data set to run analyses and to determine whether an educational technology is moving the needle on desired outcomes.

Exhibit 1. Pieces of data



StudentID	Treatment status	Reading score	Female	English learner	IEP
16574	1	465	1	1	1
19825	0	420	1	0	1
16357	0	415	0	1	0
16845	1	470	0	0	0

Dummy Variable Variable that takes the value of 0 or 1 to indicate the absence (0) or presence (1) of a specific trait or background characteristic. For example, to include gender in your data, your dummy variable would be female, with males assigned a 0 and females assigned a 1.

Effect The amount by which the average outcome of the treatment group differed from the average outcome of the comparison group. This difference is how much the technology moved the needle.

Effectiveness The extent to which something succeeds in causing the desired result or outcome. The effectiveness of an educational technology will depend on a school or district's goal for that technology. For example, one district might determine that a technology is effective if it increases student test scores by more than 10 points, whereas another district might decide that the technology is effective if it increases student test scores by any amount.

Experiment A controlled test of the effect that an intervention (for example, a reading software) has on participants' outcomes. A random assignment pilot is an experiment.

Impact See Effect

Indicator Variable See Dummy Variable

Individual Identifier A unique code, number, or ID assigned to individuals. These could be student IDs, users IDs, or school IDs. Unique values for each participant enable you to easily merge data and make information anonymous. An individual identifier is also necessary to randomly assign users and nonusers if you are conducting a random assignment pilots.

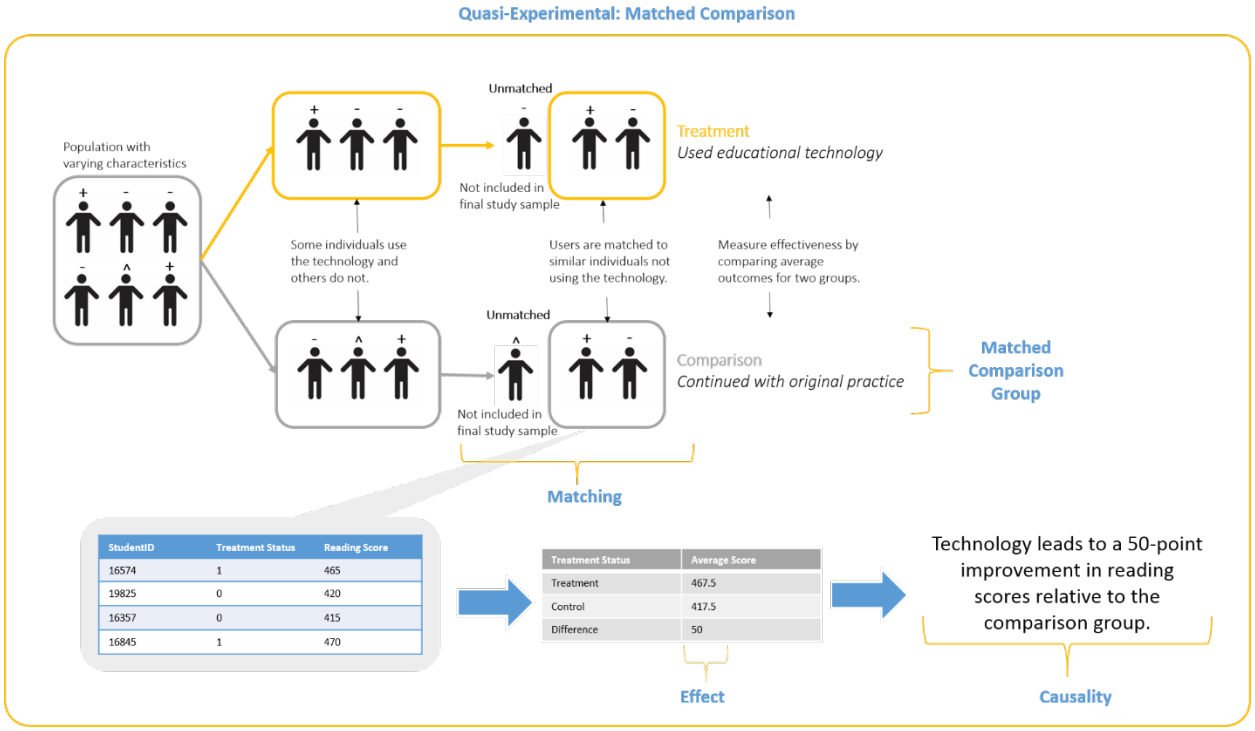
Intervention The program or educational technology being tested. In an evaluation, you investigate whether the intervention had an effect or an impact on participants' outcomes.

Intervention Group See Treatment Group

Matched Comparison Group A set of participants, in a matched comparison experiment, who are similar to the technology users but are not given access to the educational technology.

Matched Comparison Group Design An evaluation design that uses participants' pre-test and background characteristics to create two sets of similar participants. One group uses the educational technology (the treatment group) and the other group does not use the technology (the comparison group), and the outcomes of the two groups are compared. See Exhibit 2.

Exhibit 2. Pieces of a matched comparison group design



Matching The process of finding similar nonusers to compare with technology users. Only users and nonusers who are successfully matched will be compared with each other in the analysis. See Matching Overview for more information.

Missing Values A data point with no recorded information. For example, if you do not have a test score for a particular student, that would be a missing value. In a spreadsheet, these can show up as blank cells or be tagged with labels such as NA, 999, or 0. You should make sure that all missing values are labeled using NA before uploading your data to the Coach.

Numeric Type Data recorded as numbers.

Observation A single row in a data set. Each row usually represents a student, teacher, class, or school.

Observed Characteristic A variable that you can measure and that can also affect the outcome of interest. Examples of observed characteristics are individualized education program status or prior achievement.

Outcome Knowledge, skills, attitudes, or other desired benefits attained as a result of an activity. For example, student attendance or math test scores can be outcomes.

Outcome of Interest What you hope to change as a result of implementing the educational technology, such as student test scores or teacher performance.

Pilot A small-scale, short-term roll out of an educational technology. In a pilot, only a portion of potential users are introduced to the technology. The purpose is to help you learn how the technology would work on a larger scale before investing the time and resources in a full roll out.

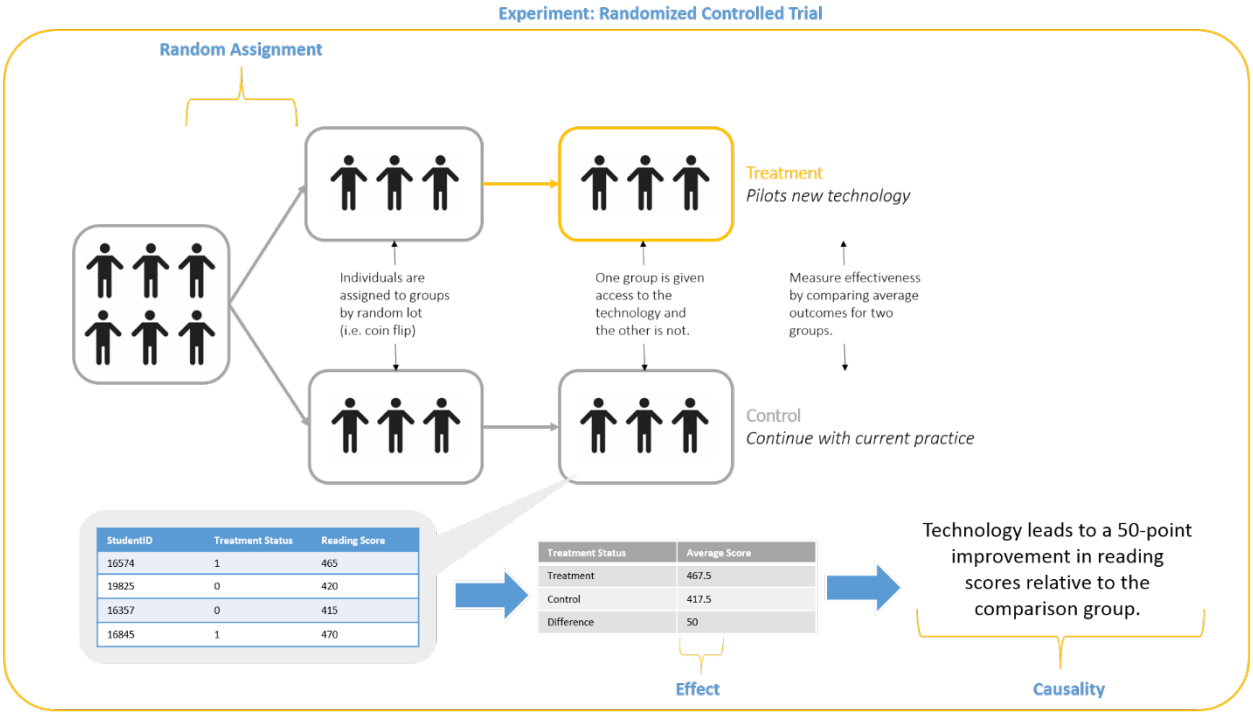
Probability The likelihood of something happening.

Quasi-Experimental A design in which groups are created through a process that is not random. For a quasi-experimental design to be rigorous, the intervention and comparison groups must be similar, demonstrating equivalence on observed characteristics before the intervention begins. A matched comparison group design is a quasi-experimental design.

Random Assignment A process by which groups of users and nonusers are formed by chance (coin flip, random number generator, and so on). When carried out correctly, random assignment results in groups that are similar on average in both observed characteristics (such as pre-test scores and gender) and unobserved characteristics (such as motivation), and any differences in outcomes between the groups are due to the intervention alone.

Randomized Controlled Trial An experiment in which subjects are assigned to the group with access to the technology and the group without access at random, for example, by tossing a coin. See Exhibit 3.

Exhibit 3. Pieces of a randomized controlled trial design



Ed Tech Rapid Cycle Evaluation Coach

Research Question An answerable question about the effect that an educational technology could have. Defining a research question is the first step in an evaluation.

Statistically Significant The likelihood that a relationship between two or more variables is caused by something other than random chance.

Treatment Group Participants who are in the group with access to the educational technology. In a randomized pilot, the treatment group is the set of participants who were randomly selected to pilot the technology. In a matched comparison design, the comparison group is usually the set of participants who actually use the technology.

Treatment Status Whether a participant belongs to the treatment (technology user) group or the comparison (nonuser) group. In a randomized controlled trial, the treatment status is the group to which the participant is randomly assigned. If a participant was assigned to the treatment group, the treatment status will always be treated, regardless of whether the participant actually uses the technology. In a matched comparison group design, the treatment status usually refers to participants who used the technology versus participants who did not use the technology. Treatment status is recorded in the data set as its own variable, with 0 indicating the participant belongs to the comparison group and 1 indicating that the participant belongs to the treatment group.

Unit of Observation Individual (could be students or teachers), school, or district for which outcomes are examined.

Unobserved Characteristic Variable that you can't see or measure that can affect the outcome of interest. For example, self-motivation is an unobserved characteristic.

Usage Data Information concerning the actions and events that users engage in while using the technology. Usage data describe all types of user interactions with a system or technology.

Variable A characteristic, assignment, or outcome that can vary or change over time or across individuals. Variables are the columns in your data set. Examples of variables include treatment status, test scores, and English learner status.

Vertically Scaled Assessment Standardized tests that are comparable across grade levels.

Frequently Asked Questions from School Districts and Educators

QUESTIONS ABOUT RAPID CYCLE EVALUATION

1. What is a rapid cycle evaluation (RCE)?

RCEs use a rigorous, quick-turnaround approach to determine whether a technology or implementation approach meets the needs of a district, school, or classroom.

2. Why is RCE important?

Conducting RCEs empowers educators to evaluate the effectiveness of educational technologies to make informed, evidence-based decisions in actionable time frames. With the implementation of the Elementary and Secondary Education Act (ESEA), as amended by the Every Student Succeeds Act (ESSA), the need for evidence-based decision making has increased significantly. The U.S. Department of Education recently released guidance on Using Evidence to Strengthen Education Investments to help state and local education agencies, schools, educators, partner organizations, and other stakeholders choose and implement successful interventions that improve outcomes for students. Conducting RCEs using the RCE Coach is one way to gather such evidence.

QUESTIONS ABOUT GETTING STARTED

3. Who should use the RCE Coach?

The RCE Coach is designed for school- or district-level staff and for educational organizations that wish to evaluate an educational technology. The RCE Coach can test the effectiveness of an educational technology that a school already uses or to set up an evaluation for a new educational technology as part of a pilot process.

Although the tool was developed for evaluating educational technologies, you can use the RCE process to evaluate other programs, interventions, and strategies.

4. How do I begin using the RCE Coach?

Visit edtechrce.org to answer a few brief questions to create an account. Review all resources and follow the step-by-step process for setting up and conducting an evaluation.

5. What if I'm not ready to start an evaluation right now?

You can access the tools and downloadable guides to learn more about the RCE process. You can find all of the tools on the Preview Tools page. If you want a better understanding of the RCE process and what you would need to conduct your own evaluation, the following tools are particularly helpful:

- RCE Overview
- What You'll Need to Get Started

6. What information does the RCE Coach provide?

The RCE Coach tells you whether an educational technology or way of implementing the technology had the desired effect. The RCE Coach provides summary statistics for the data you upload for your analysis and the results from the analysis to determine whether the technology is moving the needle. It also captures important contextual information about your district or school and how the technology was used, so that you can easily share and interpret the findings. All information generated by the RCE Coach is based on data you input into the tool.

7. What do I need to use the RCE Coach?

The key ingredients for using the RCE Coach include:

- The educational technology or implementation approach you wish to test
- A group of technology users and nonusers, or potential users and potential nonusers
- An outcome related to the technology that you can measure, such as district assessment scores
- A data set that can be uploaded to the RCE Coach
    - If you are setting up a forward-looking pilot, you have to identify the technology and data sources but will pull the information together over the course of the pilot. You will start with a list of participants who will be assigned to pilot the technology or not.
    - If you have already started using the technology and want to do a backward-looking evaluation, you might have all of this information in hand already. You will need a list of technology users and nonusers, a pre-test and outcome measure for each participant, and participant background characteristics.

8. How long should an evaluation take to complete?

On average, a typical time frame for an RCE would be three months from start to finish, but this depends on the nature of the evaluation and your research question. It's possible for some RCEs to be completed in weeks, whereas others might need more than three months. If the outcomes you're interested in take more than a few months to affect or measure, that will drive the timeline. Evaluations are also meant to be iterative, so you could go through the RCE process several times in one school year.

9. What if I care about longer-term outcomes, such as high school graduation, but want to see if things are working now?

RCE can still be useful if intermediate outcomes are necessary steps toward the longer-term outcome. In the high school graduation example, it might be the case that increasing attendance or academic performance in the short term is a good step toward improving high school graduation. You could use RCE to assess whether your technology or strategy is having an impact on attendance or achievement measures.

10. How much time will I need to invest?

The amount of time needed to conduct a successful evaluation will vary based on how far along you are in the process (for example, have you already chosen the technology?); the evaluation method; the number of stakeholders with whom you need to work; and how quickly you can collect and organize your data. If your data are already available, you could complete an evaluation in as little as a few hours.

11. Who should I involve in my RCE?

Consider including key stakeholders or decision makers and people with the skills necessary to help you complete your evaluation. Our What You'll Need to Get Started guide will help you understand what you will need to complete your evaluation. Ideally, your evaluation team will include someone with experience handling data sets. Often, people with these skills work in offices of accountability or research. However, if you do not have experienced data analysts available, the Coach provides specific instructions for preparing your data. (No prior statistics experience is necessary.) Other desirable team members include an administrative champion; staff in curriculum design, procurement, and/or instructional technology; and teachers who implement the technology.

12. How much data analysis or statistics do I need to know?

With the RCE Coach, your statistical knowledge will not be a barrier to completing a successful evaluation. The RCE Coach is designed for people with varying data analysis and statistical backgrounds. It will be helpful to have someone who is able to put together data sets involved in the process, but user-friendly guides provide all the information you need to set up your data correctly. The RCE Coach's dashboard(s) will complete the statistical portion of your analysis for you. If you have no background in these areas, you can refer to the guides provided in the RCE Coach to help you understand the findings produced for your evaluation.

13. Can I use the RCE Coach to evaluate devices?

Devices are tools for accomplishing your goals. The RCE process is optimized for evaluating concrete educational technologies, platforms, programs, procedures, and strategies that are enabled through a particular device(s). The first step in using RCE is to get specific about what you want to accomplish and the measurable outcomes that will indicate success. If you are interested in the effectiveness of a device, it is critical to think about how the device is being used to further educational goals and to evaluate specific components.

14. Can I compare two different technologies, or different implementations of the same technology?

Yes. Sometimes it is not feasible, practical, or desirable to compare technology users with nonusers. You might be more interested in learning about the best way to implement a technology, and an RCE is a great way to do that.

You can compare two different technologies. For example, you might wish to compare two different reading apps to see which has a larger impact on student reading scores. If you prefer

one app over the other, the preferred app can be designated the treatment group's technology in your evaluation. (For example, you might prefer one app because it is less expensive, or anecdotal evidence might show that it appears to be effective.) The Coach can help you test whether one technology is more effective than the other.

You can also compare two groups that use the same technology in different ways. For example, one group of students could be assigned to use a math app as homework and the other group could use the app as part of classroom lessons. The Coach can help you test whether one implementation strategy is more effective than the other.

In both cases, you will be able to tell only whether one approach was more effective than the other, not whether the technology itself was effective. Also, if you are comparing similar technologies or strategies, the two approaches could be equally effective and your results might be inconclusive.

15. What if I'm targeting use of a technology to a specific set of students? Can I still find out if the technology is effective?

To assess the effectiveness of a technology, it's critical to have a good measure of how students would have performed without it. This can be challenging when the technology is targeted to a specific set of students. For example, if you use a specific technology with all English learner (EL) students who are not reading at grade level, you would not have a comparison group that would enable you to draw strong conclusions about the product's effectiveness. However, if you have a capacity constraint—you don't have enough licenses to use the product with all EL students or you have to roll it out over time across the district—you could set up a randomized pilot that would enable you to assess effectiveness. This guide to opportunistic experiments describes situations in which this is a good option. Even when you use a technology with a specific set of students, you could use RCE to evaluate different implementations of the product or different strategies for incorporating the technology for those students. For example, you could randomly select some of the EL students to use the reading technology more frequently than others. Alternatively, you could vary the subjects in which students use the technology.

QUESTIONS ABOUT EVALUATION DESIGN

16. What is the benefit of random assignment over the other designs?

The goal of an evaluation is to determine the impact of the educational technology based on differences in outcomes between users and nonusers. To be confident that those differences are due solely to use of the technology, you will have to reduce the chance that something else is affecting the outcome. Random assignment is the best way to achieve that because it creates groups that are similar on both observed (such as pre-test scores and gender) and unobserved characteristics (such as motivation). Matching, the other design supported by the RCE Coach, enables you to create groups that are similar only on observed characteristics. Therefore, the

benefit of random assignment is that it will give you the highest level of confidence in your findings.

17. Is random assignment unfair?

You are likely implementing a technology because you assume it will benefit users. But for newer technologies, we might not know for sure how they will work in your context and with your students. The best way to learn whether the technology works is by piloting it—or offering it to a small group of users on a trial basis. In any pilot there are going to be people who test a technology first, while others will have to wait, a fair way to determine who gets the technology and who does not is through a random chance process (like flipping a coin). If the technology has an impact, more people can get access when you know that it will benefit them.

18. What questions can I answer using the RCE Coach?

The RCE Coach will enable you to answer the following question: Is this educational technology moving the needle? The RCE Coach can answer this question based on any type of user and any outcome of interest you are measuring.

19. Are there situations when RCE is not a good choice?

Yes. RCE is best suited to answering narrow questions—often looking at the effect of a specific technology or strategy—on an outcome that can be expected to be affected quickly. RCE might not be a good choice to assess the effectiveness of a district-wide initiative that has many components. However, you might be able to use RCE to evaluate whether the specific components of that initiative are achieving their goals. These results can then help you improve how you implement the initiative. In this way, RCE can play an important part in a broader culture of learning and improvement. For example, RCE could not effectively evaluate the impact of rolling out one-to-one computing district-wide, but it might be able to help you assess the effect of a new class assignment application, enabled by the one-to-one initiative, on student homework completion. For more complex or comprehensive research questions, it could be valuable to work with a research partner. The Regional Educational Laboratories or a local university can connect you with researchers in your area.

20. How large does my sample have to be? My school might be too small to use the RCE Coach.

Many factors contribute to the recommended sample size. These include whether you are interested in individuals (such as students) or groups of individuals (such as classrooms or schools); how large of an effect you expect the technology will have; and how much the characteristics of students, teachers, or schools vary. In general, a larger sample is better. The more students, teachers, or schools you have in your sample, the more likely it is that you will receive an answer to your research question (rather than inconclusive results). However, samples do not necessarily have to be huge, and having some data is better than having no data. For example, we have seen successful evaluations with as few as 30 students.

21. I am rolling out a technology to all 5th-grade students. What kind of evaluation can I run?

To conduct an evaluation you must be able to compare a group using the technology to a similar group not using the technology. If all students in a specific grade are using the technology, you would not have any nonusers with which to compare them. You could consider rolling out the technology in phases; that way you would have a short period during which you have users and nonusers. You could also implement the technology in different ways, and compare outcomes to support one implementation method over the other. For example, you could have some students use the technology for 60 minutes per week and others use it for 120 minutes per week. Depending on how you choose who uses the technology for a specific amount of time, you could use the randomized pilot or matched comparison pathways that the Coach offers.

22. If all of my students are using the technology, can I use usage data to categorize my treatment and control groups—by designating low-usage users as my control group?

Though this gives you some information, it does not provide the convincing evidence you need to conclude that the educational technology is leading to the observed outcomes. This approach, however, could be a good starting point for developing further ideas to test in a subsequent pilot. This is because assigning students to groups based on their technology use increases the likelihood that unobserved differences in students affect outcomes. Other factors could lead some students to use the technology less, and those factors might also affect outcomes. For example, for a technology that students use at home, poor technology or limited internet access could lead to lower use. Those factors could also be associated with poorer student outcomes. If that is the case, you wouldn't be able to determine whether it was low usage or technology and internet access in general that caused the observed outcomes.

23. Does the RCE Coach enable me to analyze the effects of the technology on subgroups, such as boys or English learners?

You can analyze subgroups using the Coach; however, to do this in the current version of the Coach you will have to go through each step of the Coach for each subgroup of interest. You will need a data set specific to each subgroup to analyze outcomes by subgroup. (For example, to analyze effects by gender, you will need one data set that has only male students and another data set that has only female students.)

24. Can I look at multiple outcomes?

The RCE Coach is currently designed to analyze one outcome at a time. However, you can easily select one outcome, get results, produce a findings brief and then repeat that process for a second outcome. If this is a feature you would like to see, please let us know at EdTechRCE@mathematica-mpr.com.

25. Do I need to use the same pre-test and post-test to measure the outcome?

No, you do not need to use the same pre-test and post-test. The effect of the technology is based on the comparison of two groups, not on pre-test versus post-test measures. The Coach uses the pre-test scores to create similar groups (in the case of a matched comparison design) or to check

that both groups are equivalent (in the case of a randomized pilot). The Coach also accounts for pre-test scores in the final analysis. Whenever possible, the pre-test should measure the same thing as the post-test. For example, if you are interested in student math achievement, the Coach recommends that you use a math test for the pre-test and the post-test, but it does not have to be the same math test.

26. One school (or one class) in my district is using a technology. Can I compare the students in this school (or class) with other students in the district?

To conduct an evaluation you have to be able to compare a group using the technology with a similar group not using the technology. If only one school or class is using the technology, it will be impossible to disentangle the effect of the technology from the effect of characteristics that are unique to that group (such as the quality of the school or teacher). However, if you have data from a situation like this, you can use it to start to build evidence. Comparing results from one school or classroom with another group can provide suggestive evidence. To build more confidence in those results, we recommend that you set up a forward-looking pilot that includes more than one classroom or school in each group.

27. Can I compare 7th-grade students who use the technology this year to 7th-grade students who did not use the technology last year?

The RCE Coach does not currently support these types of comparisons for causal analysis. Environments change over time. If you compare 7th graders one year to 7th graders the following year, any differences you see could be due to differences in the students, or to other programs, policies, practices, or contextual factors that changed from one year to the next. However, if you have data from a situation such as this, you can use it to start to build evidence. Comparing results from across years can provide suggestive evidence. To build more confidence in those results, we recommend that you set up a forward-looking pilot that compares students from the same year.

QUESTIONS ABOUT INFORMATION SHARING

28. How can I share the results of my evaluations?

The RCE Coach enables you to generate a portable document format (PDF) file of your findings or print the results from any web browser. We recommend that you follow your existing protocols for sharing results in your district or organization.

29. How is student information handled in the RCE Coach?

The RCE Coach does not require personally identifiable student information in the data set you will need to upload (called user data files in the Privacy Policy) to use the tool. The RCE Coach provides guidelines on how to prepare and upload data sets (user data files), including specific measures for de-identifying students and avoiding personally identifiable information (PII). However, the RCE Coach cannot prevent you from uploading data sets that include PII. As you prepare to create your evaluation report or brief, the Coach will ask you to enter information (this is called user input in

the Privacy Policy). The RCE Coach will provide tips along the way to help you avoid inadvertently entering PII. Finally, the evaluation reports or briefs that the tool produces (called user output in the Privacy Policy) will contain only student information that is sufficiently aggregated so that it is nonidentifiable.

30. How does the RCE Coach handle privacy and Federal Educational Records Privacy Act of 1974 (FERPA)?

The RCE Coach does not require personally identifiable student information to use the tool or create evaluation reports or briefs that contain PII. When you carefully follow the instructions provided by the tool, your use of the tool complies with FERPA. The RCE Coach includes a Privacy Policy and Terms of Use for users to acknowledge when setting up an account. Please visit our Privacy Policy and Terms of Use for more information.

31. What happens to the data files that I upload into the RCE Coach?

The data sets or data files you upload (called user data files in the Privacy Policy) are stored exclusively in a temporary cache. They are not stored permanently anywhere in the tool. The data contained in your user data files will be used only to produce the analysis, brief, or report that you requested. The user data files are not used for any other purpose. When you are finished and leave that page, your user data files are permanently erased from the temporary cache.

32. What happens to reports after the evaluation? Who owns the information?

You always own all of the information that you put into the RCE Coach and all of the information that comes out of your use of the RCE Coach. However, the terms of use and Privacy Policy provide that you give Mathematica Policy Research the right to use, so long as it is aggregated and or de-identified, the information you entered into the RCE Coach (the Privacy Policy calls this your user input) and the evaluation reports and briefs created by RCE Coach (the Privacy Policy calls this your user output) for educational, academic, and research purposes and to help other users or to promote research and development of technology like the Ed Tech RCE Coach. Your user input (information you entered) and user output (your evaluation reports and briefs) are stored permanently on Amazon Web Services under your RCE Coach account. Remember, your data set or data files you uploaded (the Privacy Policy calls these your user data files) are not stored anywhere, but are deleted when you finish using the RCE Coach. Also, your evaluation reports and briefs could be reviewed by the What Works Clearinghouse (WWC) if the evaluation meets its screening criteria and could be included in the WWC | Find What Works database. Please visit our Privacy Policy and Terms of Use for more details.

QUESTIONS ABOUT GETTING ADDITIONAL ASSISTANCE

33. I want to know what I am getting into. Is there an overview of all of the steps?

Check out What You'll Need to Get Started for an overview of the skills and data you will need to conduct a successful evaluation. You can also find a list of all of the tools under Preview Tools.

34. A lot of terms are unfamiliar to me. Is there a printable glossary that I can use as a reference?

Yes. The RCE Coach uses technical terms only when absolutely necessary. In these cases, there is often a definition that you can view by placing your cursor over the word. For more extensive definitions, you can view our printable glossary. The technical vocabulary is commonly used in educational research and becoming familiar with these terms will help you understand other research as well.

35. I don't have the answers to all of the questions.

a. Are there some responses that are required and others that are not?

Yes. Answering all of the questions will enable you to produce the most complete findings brief, but only some questions are required to complete the analysis. Some tools will be locked (and allow you only to peek at the contents) until you answer required questions. However, if you skip many questions, it will likely be difficult to interpret and share your findings.

b. Can I skip a question and return later?

Yes. The RCE Coach is designed so you can move through and answer questions as you have answers.

c. Will the RCE Coach remember which questions still have to be completed?

Yes. The RCE Coach tracks the questions you have completed and prompts you for key information when it is required.

36. Who do I contact if I need help using the tool?

You can leave questions and your contact information in the Feedback tab on the right margin of each page and someone from the RCE project team will contact you. You can also periodically check the frequently asked questions for updated responses.

37. How can I pilot the RCE Coach and receive technical assistance?

The RCE Coach is freely available for anyone to use. Resources are available to support a number of districts during 2017 in conducting RCEs with the RCE Coach. If you are interested in receiving technical assistance to help you get started, please complete the Ed Tech RCE Pilot Project Form to share some preliminary information with the RCE project team. We will release a request for proposals in the near future for winter and spring pilots. We are particularly interested in forward-looking pilots that can take advantage of the new features released in January 2017. In addition, we would like to support districts that wish to test educational technologies with a variety of goals, such as teacher productivity and teacher professional development, as well as student achievement.

QUESTIONS ABOUT BEING A PILOT PARTNER

38. What are the criteria for being selected as a pilot partner?

Your district would be a strong candidate if you:

- Are considering piloting a new technology and are open to using a lottery-based approach to assign use of the technology
- Have implemented a technology in some schools or classes but not others, providing opportunities to compare outcomes for similar students or teachers who have access to the technology with others who don't
- Are interested in testing different approaches to implementing a technology, such as different teacher training programs or different communication strategies

39. What kinds of software applications are being considered?

We are interested in evaluating technologies that are relatively untested and can provide high quality data on whatever outcomes the technology aims to impact. Technologies can be aimed at student achievement, social-emotional learning, teacher professional development, teacher productivity, or another district goal.

40. What does participation entail?

Participating districts will work with our research partners to conduct one or more technology evaluations for your school(s). We will seek your input on the toolkit being developed to support RCEs, giving you an opportunity to help shape this important national initiative. Your staff may participate in teleconferences with developers and build capacity to obtain and interpret system data generated by technologies. We will summarize findings from the RCEs in short reports and make those reports available to the public.

FREQUENTLY ASKED QUESTIONS FROM EDUCATIONAL TECHNOLOGY COMPANIES

41. Can my company use the RCE Coach to evaluate our educational technology product?

Yes, the RCE Coach is free and available for anyone to use. However, it is specifically designed for educators (school districts, educational leaders, and individual educators). We encourage companies to partner with school clients that are willing to evaluate your product using the RCE Coach.

42. As an educational technology company, what can I expect from school clients that use the RCE Coach to evaluate my product?

School clients will likely request specific usage data that they need to create the data sets that they will upload into the tool. Based on results of evaluations, some school clients might suggest modifications to your technology to improve intended outcomes.

43. What schools and companies are currently participating?

We are unable to release this information publicly. However, our Shared Evaluations section identifies some participating schools, school districts, and products.

44. When and where are evaluations published for public viewing?

Evaluations are available for viewing on the Shared Evaluations page and more evaluations will be added as they are completed. Users can opt in or out of sharing evaluations publicly.

45. Why should vendors participate in the RCE pilot process?

- The field is moving in the direction of evidence-based decision making.
- The ESEA, as amended by ESSA, is putting increased emphasis on evidence, and an increasing number of schools will be making evidence-based decisions about adopting educational technology.
- The RCE process will help you to implement more efficient pilots.
- You can give us feedback to improve the tool, implementation, and pilot process.
- There is no cost for you to participate.
- You will receive valuable information that you can use to inform product development.